

Review

Automated Skin Lesion and Cancer Detection Using Computer Vision: A Comprehensive Review

Bhagyashri S. Sonune ¹, Udayakumar Ramanathan ¹, Dhiraj P. Tulaskar ², Shon G. Nemane ²,
Madhusudan B. Kulkarni ^{3,*}, Prakash Rewatkar ⁴ and Manish Bhaiyya ^{2,*}

¹ Department of Computer Science and Information Technology, Kalinga University, Raipur 492101, CG, India

² Department of Electronics and Telecommunication Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon 444203, MH, India

³ Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, KA, India

⁴ Department of Mechanical Engineering, Israel Institute of Technology, Haifa 3200003, Israel

* Correspondence: madhusudan.kulkarni@manipal.edu (M.B.K.); bhaiyya.manish@gmail.com (M.B.)

Abstract

Detection of skin cancer has become an increasingly prevalent health issue for which there is a need for reliable methods of detection to improve patient outcomes, minimize delays in diagnosis, and ensure clinical referral. Advances in computer vision and artificial intelligence have allowed automatic analyses of dermoscopic, clinical, and smartphone imaging of skin lesions. While numerous models have been found to perform very well on the basis of curated benchmark datasets, their accuracy in practical settings still needs to be evaluated. This review critically evaluates the literature from 2015 to 2025. Rather than considering the Dice, Jaccard, AUC, sensitivity, and specificity metrics on an absolute basis, this review evaluates them relative to dataset quality, validation process, external testing, statistical analysis, and risk of bias. Key areas of focus include dataset imbalance, lack of coverage of darker skin, poor external validation, explainability, uncertainty quantification, and barriers to clinical adoption. This review provides valuable insights for researchers, practitioners, dataset producers, and healthcare providers involved in developing AI-enabled dermatology tools.

Keywords: skin lesion; cancer; computer vision; artificial intelligence (AI); detection; dermatology

1. Introduction

Skin cancer comprises one of the most prevalent and clinically significant types of cancer that exist in the world today, encompassing not only melanoma but also non-melanoma forms of skin cancer such as basal cell carcinoma and squamous cell carcinoma [1–4]. While the latter types may be more prevalent, the former, namely melanoma, poses great clinical concern due to its higher tendency for metastasis and associated risk of mortality when left undiagnosed [5–8]. The early identification of skin cancer is crucial not only for the improvement of patient survival but also for avoiding unnecessary biopsy procedures and for timely referral to specialists. Typically, the process of skin cancer diagnosis involves visual inspection and then dermoscopy, providing a better visibility of pigmentation structure, vasculature, boundaries of lesions, and other features that cannot be observed visually (see Figure 1). However, the effectiveness of the procedure will depend on many factors, such as the physician's experience and level of training, quality of images, lesion morphology,



Academic Editor: Andrea Cataldo

Received: 21 May 2026

Revised: 2 July 2026

Accepted: 3 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

anatomical site, patient skin pigmentation, and availability of dermoscopy or dermatology services [9–11].

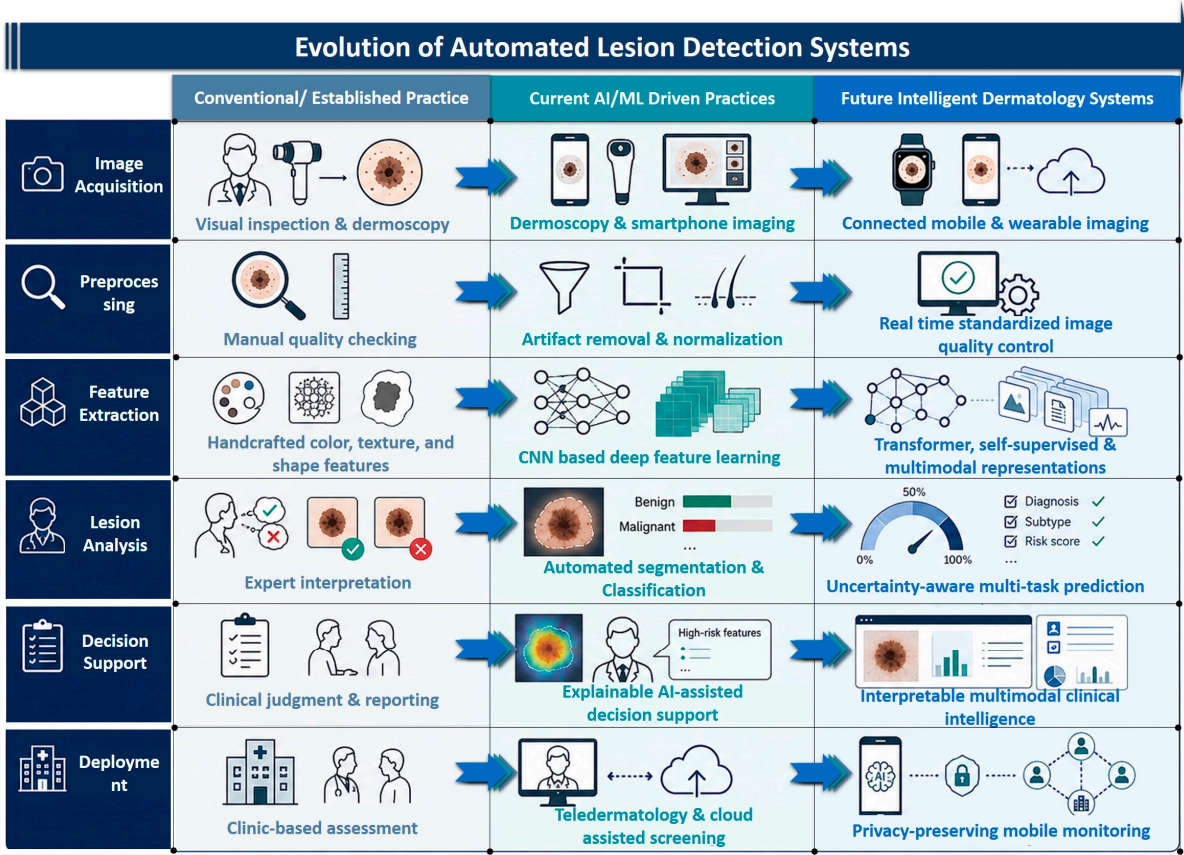


Figure 1. Evolution of skin lesion detection systems from conventional methods to AI-ML-empowered frameworks.

There is a need for computer-aided diagnosis tools that will help eliminate subjectivity and provide objective information for the assessment of skin lesions. Previously, computer-aided diagnosis was based on engineered features such as asymmetry, border irregularity, pigmentation differences, diameter, texture, and shape characteristics [12–15]. Such methods were very useful in building a foundation for lesion characterization using computers; however, they were prone to various lighting effects, interference by hairs, image acquisition effects, segmentation errors, and low diversity of the datasets used. The fast development in machine learning and deep learning has dramatically changed the way skin lesions are assessed. There are many computer algorithms available that can analyze dermoscopic, clinical, and mobile phone pictures automatically using techniques such as convolutional neural networks, transfer learning, ensemble modeling, attention mechanisms, and transformers. Publicly available databases like ISIC, HAM10000, PH2, Derm7pt, and SD-198 have sped up the development of these algorithms through benchmarking [16–21].

However, despite all the progress, there are still problems with translating the use of AI-based algorithms into clinical practice. There are many reports about the good internal performance of such algorithms, where Dice and Jaccard scores for segmentation and AUC, accuracy, sensitivity, and specificity values for classification show promising results [22–27]. Nevertheless, these metrics are usually obtained using curated datasets, internal validation splits or image-level partitions that do not cover real-world clinical diversity. The differences in imaging devices used, acquisition protocol, lesion prevalence, annotation quality, class balance and demography could affect the performance of the models negatively

when these algorithms are used on separate datasets [28,29]. Furthermore, the lack of representation of dark Fitzpatrick skin types, rare lesion sub-types, clinical images without dermoscopy and smartphone images is the problem of many public datasets [30–33].

Several review articles have summarized progress in skin cancer detection using machine learning, deep learning, convolutional neural networks, patient metadata, and clinical deployment perspectives. Earlier reviews mainly focused on conventional image-processing pipelines or CNN-based classification, while more recent systematic reviews have examined clinician–AI comparison, primary-care applicability, commercial dermatoscopic systems, Asian populations, and melanoma diagnosis or prognosis [34–41]. However, most previous reviews either emphasized classification alone, focused on dermoscopic images, discussed deep learning methods narratively, or did not integrate segmentation, classification, generative augmentation, multimodal fusion, explainable AI, fairness, external validation, and risk-of-bias appraisal within a single framework. Therefore, as summarized in Table 1, there remains a need for a more integrated and critically appraised review that evaluates not only what performance values are reported, but also how reliable those values are in terms of dataset quality, validation design, external testing, and clinical applicability.

Table 1. Comparative analysis of existing review articles and novelty of the present.

Published Review	Main Focus	Key Contribution	Limitation/Gap	How the Present Review Is Different
Mehta and Shah, 2016 [42]	Computer-aided skin cancer diagnosis	Summarized classical CAD steps such as preprocessing, segmentation, feature extraction, and classification.	Mainly focused on traditional image-processing pipelines before the deep-learning expansion.	Extends beyond classical CAD to CNNs, transformers, GANs, multimodal fusion, explainability, and clinical translation.
Adegun and Viriri, 2021 [4]	Deep learning for skin lesion analysis and melanoma detection	Provided a survey of deep-learning techniques used for skin lesion images.	Primarily technique-oriented; limited formal evidence-strength appraisal.	Adds critical appraisal of reported metrics using validation quality, external testing, and risk-of-bias considerations.
Dildar et al., 2021 [43]	Deep learning for skin cancer detection	Reviewed deep-learning-based methods for early diagnosis of skin cancer.	Strong classification emphasis; limited integration of multimodal and generative methods.	Integrates segmentation, classification, GAN-based augmentation, multimodal learning, and deployment challenges.
Höhn et al., 2021 [44]	CNNs with patient data	Systematically reviewed integration of patient metadata into CNN-based skin cancer classification.	Focused specifically on patient-data integration.	Discusses multimodal fusion as one component within a broader AI pipeline including segmentation, classification, and generative models.
Haggenmüller et al., 2021 [35]	CNN classification compared with human experts	Reviewed CNN-based skin cancer classification studies involving human experts.	Mainly clinician-comparison oriented.	Places classification performance within broader concerns of dataset bias, external validation, fairness, and interpretability.
Jones et al., 2022 [45]	AI/ML for early skin cancer detection in primary and community care	Evaluated AI/ML algorithms for early diagnosis in community and primary-care settings.	Strong clinical-setting focus but less emphasis on technical evolution across segmentation, GANs, transformers, and multimodal AI.	Connects technical model development with real-world clinical-readiness barriers.
Wu et al., 2022 [34]	Skin cancer classification with deep learning	Reviewed recent developments in dermoscopic skin lesion classification.	Mostly classification-centered and dermoscopy-focused.	Expands scope to segmentation, generative augmentation, multimodal fusion, explainability, and evidence strength.
Debelee et al., 2023 [46]	Skin lesion classification, segmentation, and detection	Surveyed recent machine-learning methods for skin lesion analysis.	Broad method summary with limited formal risk-of-bias interpretation.	Adds structured quality appraisal and interprets metrics according to validation rigor.

Table 1. Cont.

Published Review	Main Focus	Key Contribution	Limitation/Gap	How the Present Review Is Different
Nazari and Garcia, 2023 [47]	Skin cancer detection using clinical images	Focused on standard clinical images rather than only dermoscopy.	Does not comprehensively cover dermoscopic benchmarks, generative models, and multimodal fusion.	Combines dermoscopic, clinical, and smartphone-image perspectives with AI-model comparison.
Salinas et al., 2024 [48]	AI versus clinicians for skin cancer diagnosis	Systematic review and meta-analysis comparing AI and clinician performance.	Focused mainly on diagnostic accuracy comparison.	Goes beyond AI-versus-human comparison to assess methodological reliability, bias, and deployment readiness.
Ang et al., 2025 [49]	AI skin cancer detection in Asian populations	Addressed population-specific concerns and adaptation of AI models to Asian cohorts.	Region/population-specific scope.	Includes fairness and subgroup limitations as part of a broader global review framework.
Karimzadghagh et al., 2026 [50]	Umbrella review of systematic reviews and meta-analyses	Synthesized evidence from prior reviews and meta-analyses.	Umbrella-level focus; less detailed technical synthesis of segmentation, GANs, transformers, and multimodal methods.	Provides method-level synthesis with detailed discussion of AI pipelines, metrics, validation, and clinical translation.

Novelties and Contributions of This Review

This review helps fill the abovementioned research gap through a systematic and critical orientation to synthesize skin lesion and cancer detection via computer vision and artificial intelligence approaches between 2015 and 2025. As opposed to prior reviews, which mainly considered classification accuracy, this review encompasses four related fields of study: lesion segmentation, image classification, data generation, and multimodal fusion. Furthermore, the review looks at emerging research trends, including transformer models, self-supervised learning, interpretable AI, uncertainty-aware prediction, and privacy-preserving federated learning. In total, 125 primary studies were reviewed from among 842 screened records through the PRISMA-oriented process detailed in the Methodology section.

The key contribution of this systematic review lies in the interpretation of Dice, Jaccard, AUC, sensitivity, and specificity scores in relation to the quality of the methodology as opposed to their direct comparability as outcome measures. In order to improve the systematic aspect of this manuscript, the review makes use of an appraisal perspective of PROBAST-AI/TRIPOD-AI type, taking into account representativeness of the dataset, quality of annotation, control of data leakage, validation approach, external validation, subgroup or fairness analysis, interpretability, reproducibility, and clinical significance.

This review is a consolidation of computer vision techniques used for the automatic detection of skin lesions and cancers. In addition to progress in the field, this review provides evidence of the challenges that still need to be overcome before such AI technology can be viewed as reliable and equitable.

2. Review Methodology

This review was conducted using a systematic and reproducible methodology designed to identify, screen, extract, and critically appraise studies on automated skin lesion and skin cancer detection using computer vision and artificial intelligence. The review process followed the PRISMA 2020 framework and included database searching, duplicate removal, title and abstract screening, full-text eligibility assessment, structured data extraction, and risk-of-bias/quality appraisal. In response to the methodological heterogeneity of artificial-intelligence-based diagnostic studies, the synthesis was not limited to reporting numerical performance values; instead, reported Dice, Jaccard, AUC, sensitivity, and

specificity values were interpreted in relation to validation design, dataset characteristics, annotation quality, leakage control, and external validation.

2.1. Search Strategy and Information Sources

A comprehensive literature search was conducted across five major scientific databases: PubMed, IEEE Xplore, Scopus, Web of Science, and arXiv. The search covered studies published between 1 January 2015 and 31 March 2025. This time window was selected because the period after 2015 corresponds to the rapid expansion of deep-learning-based skin lesion analysis, particularly following the widespread adoption of convolutional neural networks, public dermoscopy benchmarks, and challenge datasets such as ISIC.

The search strategy combined lesion-related, imaging-related, and machine-learning-related terms. Representative search terms included: “skin lesion”, “skin cancer”, “melanoma”, “non-melanoma skin cancer”, “dermoscopy”, “clinical image”, “smartphone image”, “computer vision”, “artificial intelligence”, “machine learning”, “deep learning”, “convolutional neural network”, “CNN”, “transformer”, “segmentation”, “classification”, “GAN”, “generative model”, “multimodal fusion”, “metadata”, “explainable AI”, and “federated learning”. Boolean operators were used to combine these terms according to the syntax requirements of each database.

In addition to database searching, manual searches were performed using known benchmark resources and dataset-related literature, including the ISIC Archive, HAM10000, PH2, Derm7pt, and SD-198 datasets. The reference lists of eligible articles and relevant review papers were also screened to identify additional primary studies that may not have been captured through database searching.

2.2. Eligibility Criteria

Studies were included if they met the following criteria:

- (i) They were primary research articles;
- (ii) They addressed automated skin lesion or skin cancer analysis using dermoscopic, clinical, or smartphone-based images;
- (iii) They investigated at least one of the following tasks: lesion segmentation, image classification, generative augmentation, multimodal fusion, explainability, or deployment-oriented AI methods;
- (iv) They reported quantitative performance metrics such as THE Dice coefficient, Jaccard index, AUC, accuracy, sensitivity, specificity, precision, F1-score, or related diagnostic indicators.

Studies were excluded if they were review articles, editorials, commentaries, conference abstracts without sufficient methodological detail, non-image-based studies, non-skin-cancer studies, duplicate publications, or studies published before 2015. Studies using private datasets were excluded when the dataset origin, annotation protocol, class distribution, or evaluation design was insufficiently described. Studies that did not report quantitative outcomes or did not provide enough methodological information to assess model development and validation quality were also excluded.

2.3. Study Selection Process

The identification of all articles from the literature search was performed using a referencing procedure, during which any duplicate records were discarded. The next step of screening involved the removal of clearly irrelevant articles at the level of titles and abstracts by two independent reviewers. Finally, those potentially relevant articles that remained were independently assessed by two reviewers against predefined inclusion and exclusion criteria.

Methodological rigor was enhanced through inter-reviewer reliability testing via Cohen's Kappa. Inter-reliability testing via Cohen's Kappa produced high inter-reliability scores between the two reviewers for both screenings. The inter-reliability score via Cohen's Kappa was 0.86 for the screening of titles and abstracts, while that of full texts was 0.82. Discrepancies were sorted out through discussions and if not settled, a third reviewer decided the issue.

The first phase of the literature search generated 842 citations from searches through databases and other sources. After the elimination of duplicates and screening for eligibility, 125 primary studies were selected for inclusion in this systematic review. The selection of studies is illustrated by the PRISMA flow diagram in Figure 2 below. The studies were categorized based on their respective methodological approach as follows:

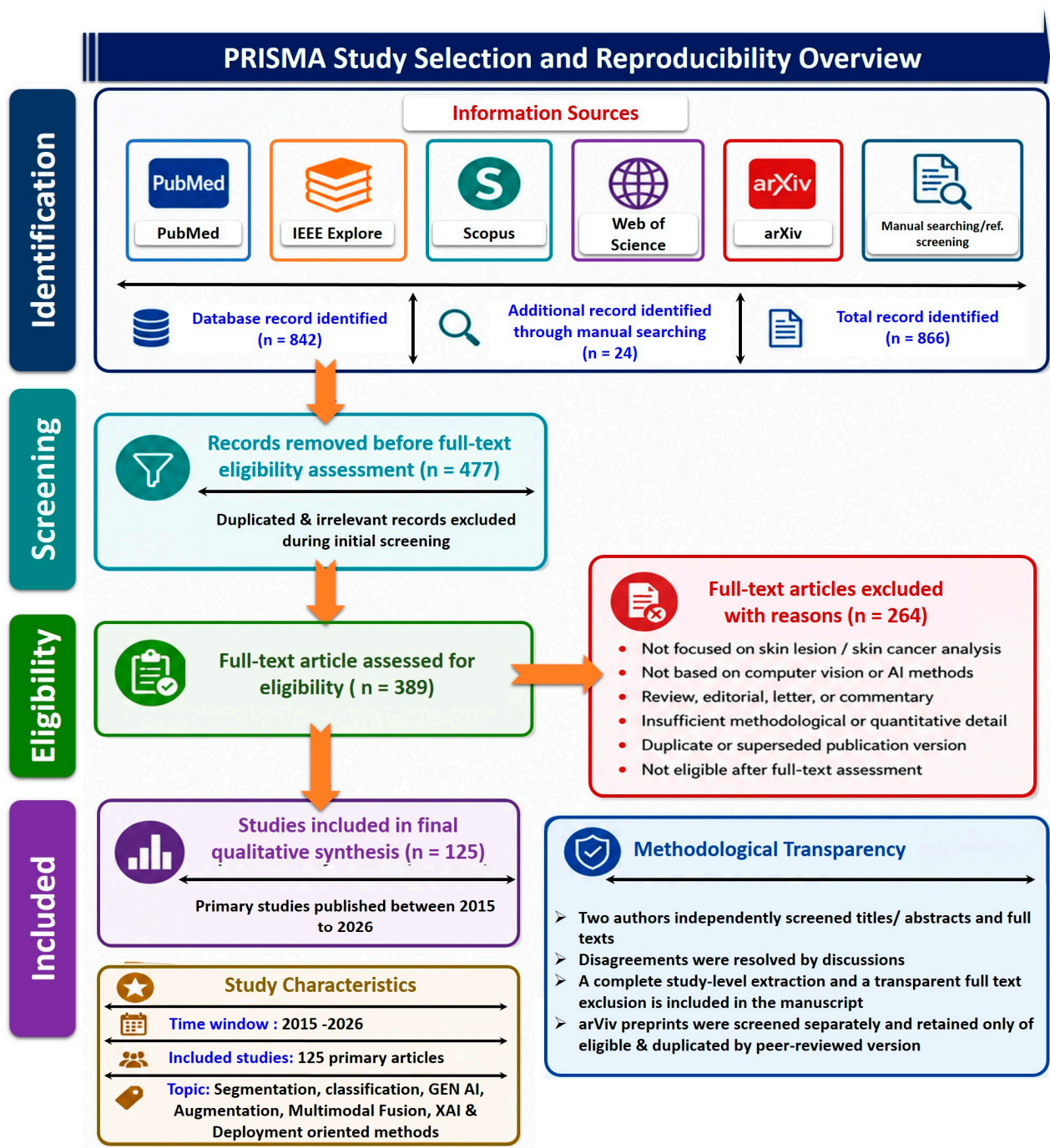


Figure 2. PRISMA 2020 study-selection and reproducibility workflow for the systematic review.

2.4. Data Extraction

Data were extracted using a standardized template to ensure consistency across studies. The extracted items included publication year, study objective, imaging modality, dataset name, dataset size, lesion classes, model architecture, preprocessing methods, augmentation strategy, segmentation or classification task, training–validation–test data split, validation approach, reported performance metrics, external validation status, interpretability method, and major limitations reported by the authors.

For segmentation studies, extracted metrics included the Dice coefficient, Jaccard index, sensitivity, specificity, boundary-related indicators, and annotation details where available. For classification studies, extracted metrics included AUC, accuracy, sensitivity, specificity, precision, F1-score, and class-wise performance where available. For generative augmentation studies, extracted information included GAN type, conditioning strategy, synthetic-image evaluation, effect on downstream classification or segmentation, and whether synthetic data improved external validation or only internal performance. For multimodal studies, extracted information included image features, clinical metadata, fusion strategy, missing-data handling, and performance comparison between image-only and multimodal models.

2.5. Quality Assessment and Risk-of-Bias Appraisal

To strengthen the systematic-review characteristic of this work, a formal quality-assessment and risk-of-bias appraisal was performed for the included AI diagnostic studies. The appraisal framework was adapted from PROBAST+AI and TRIPOD+AI-style principles. PROBAST+AI is intended for evaluating quality, risk of bias, and applicability of prediction models using regression or artificial intelligence methods, while TRIPOD+AI provides updated reporting guidance for prediction-model studies that use regression or machine-learning approaches.

Each study was assessed across the following domains: data source and population representativeness, reference standard or annotation quality, dataset splitting and leakage control, preprocessing transparency, augmentation strategy, model-development procedure, internal validation, external validation, performance reporting, fairness/subgroup analysis, interpretability, uncertainty estimation, reproducibility, and clinical applicability.

Each domain was rated as low concern, some concern, or great concern. An overall evidence-strength judgment was then assigned to each study based on the combined methodological profile. Studies with a transparent dataset composition, patient-level data splitting, reliable annotation or reference standards, complete performance reporting, and independent external validation were considered to provide stronger evidence. Studies based only on internal validation, single-dataset testing, unclear splitting strategy, incomplete metric reporting, or limited demographic diversity were considered to provide moderate or weaker evidence, even if their reported Dice, Jaccard, or AUC values were high.

This appraisal was used to avoid over-interpreting numerical performance metrics. For example, a segmentation model reporting a high Dice score on an internal ISIC test split was not interpreted as having the same strength of evidence as a model evaluated on an independent external dataset. Similarly, a classification model reporting high AUC without sensitivity, specificity, confidence intervals, calibration, or subgroup analysis was considered less clinically reliable than a model reporting complete diagnostic metrics and external validation. Supplementary Materials were added to improve transparency and reproducibility of the search, screening, extraction, and quality-assessment process.

2.6. Appraisal Domains Used for Evidence Evaluation

Table 2 presents the quality assessment framework employed to evaluate the methodological quality and risk of bias of the included AI-based diagnostic studies.

Table 2. Quality-assessment framework used for included AI diagnostic studies.

Domain	Key Items Assessed	Relevance to Evidence Strength
Data source and representativeness	Dataset size, lesion diversity, class balance, imaging source, skin-tone diversity, single-center or multi-center origin	Determines whether the reported model performance is likely to generalize beyond the training dataset
Reference standard/annotation quality	Histopathology, expert diagnosis, consensus labels, segmentation-mask quality, annotation protocol	Weak labels or inconsistent masks can distort Dice, Jaccard, sensitivity, and specificity
Dataset splitting and leakage control	Patient-level separation, duplicate removal, train-validation-test data split, cross-validation strategy	Image-level splitting or duplicate leakage may inflate performance metrics
Preprocessing transparency	Hair removal, resizing, color normalization, artifact removal, contrast enhancement	Poorly described preprocessing reduces reproducibility
Augmentation strategy	Conventional augmentation, oversampling, GAN-based synthesis, minority-class balancing	Augmentation can improve internal metrics but may not improve external generalization
Model-development strategy	Architecture, transfer learning, hyperparameter tuning, ensemble method, threshold selection	Excessive tuning on benchmark datasets may overestimate real-world performance
Validation design	Internal validation, cross-dataset validation, external validation, multi-site testing	External validation provides stronger evidence than single-dataset internal testing
Performance reporting	Dice, Jaccard, AUC, accuracy, sensitivity, specificity, confidence intervals, calibration	Complete reporting allows fair comparison across studies
Fairness and subgroup analysis	Fitzpatrick skin type, age, sex, anatomical site, rare lesion classes	Lack of subgroup analysis limits clinical reliability and equity
Interpretability and uncertainty	Grad-CAM, saliency maps, uncertainty estimation, calibration curves	Supports clinical trust but does not replace external validation
Reproducibility	Public code, model details, dataset availability, parameter reporting	Determines whether results can be independently verified
Clinical applicability	Inference time, computational cost, workflow integration, deployment setting	High benchmark performance may not translate to practical clinical use

2.7. Evidence-Strength Classification

Based on the risk-of-bias appraisal, the included studies were interpreted using three evidence-strength categories: *Stronger evidence*: These studies have clearly described datasets, reliable reference standards, patient-level splitting, complete metric reporting, and independent external or cross-dataset validation. These studies provide more reliable support for claims of model generalizability and clinical relevance.

Moderate-quality evidence: These studies have been developed in an openly accountable manner and validated internally, but not externally, or where the analysis of subgroups is partially complete, or where diagnostic criteria have only been partially reported.

Poorer-quality papers: These papers have ambiguous data partitioning, potential leakage, inadequate sample sizes or imbalance, lack of thorough reporting, lack of external validation, limited annotation information, and/or results obtained from internal test sets that have been overly curated. High Dice, Jaccard, AUC, sensitivity, or specificity results from such papers were not considered strong indicators of clinical readiness.

2.8. Data Synthesis

Because the included studies differed substantially in dataset composition, lesion categories, imaging modalities, preprocessing methods, model architectures, and validation protocols, a formal meta-analysis was not performed. Instead, the findings were synthesized qualitatively and comparatively. Performance values were summarized by methodological category, including segmentation, classification, generative augmentation, and multimodal fusion.

The reported Dice and Jaccard values for segmentation studies were interpreted together with annotation quality, lesion-boundary definition, and external-validation status. Similarly, AUC, sensitivity, and specificity values for classification studies were interpreted in relation to class imbalance, threshold selection, validation design, and dataset representativeness. For generative models, improvement in internal classification or segmentation performance was interpreted cautiously unless supported by independent testing. For multimodal studies, performance gains were evaluated in relation to metadata completeness, fusion strategy, and reproducibility.

Hence, the values provided for performance in this review cannot be seen as interchangeable across different studies. They are, rather, evidence that is shaped by the qualities of the dataset used, by the validation process, and by the risk of bias. In this way, the review can go further than just providing a narrative description of the evidence for AI performance in automated skin lesion and cancer detection.

3. Background, Clinical Context, and Dataset Characteristics

Skin cancers include both melanomas and non-melanoma skin cancers, like basal cell carcinoma and squamous cell carcinoma. Even though non-melanoma skin cancers are more prevalent than melanomas, the latter are clinically more aggressive due to their tendency to metastasize and high probability of fatalities in cases of late diagnosis [51–53]. Therefore, early detection is critical for increasing patients' survival rates, minimizing unnecessary biopsies, and ensuring timely referrals to specialized dermatologists. As part of the usual clinical procedures, the initial step in the diagnostic process typically involves examination through the naked eye, with dermoscopy being employed when possible. Using dermoscopy allows an improved view to be obtained of the subcutaneous pigmentation network, vascular components, lesion margins, and morphology not visible through conventional examination of the skin. Yet, the efficacy of dermoscopy in skin cancer diagnosis greatly depends on a variety of factors [54–56].

A computer-aided diagnosis system has been designed in order to minimize diagnostic variance and ensure objectivity in skin lesion analysis. Previous computer-aided diagnosis systems were mostly based on hand-crafted features that included asymmetry, irregular borders, color heterogeneity, texture, size, and shape features [57–60]. This approach formed the basis of automated skin lesion analysis, but it was prone to changes in illumination, hair artifacts, noise, segmentation error, and hardware differences. The development of deep learning, especially the use of convolutional neural networks and transfer-learning approaches, led to changes in the field of automated skin lesion analysis from the manual engineering of features to automated feature learning [61–64].

The clinical value of AI-based skin lesion analysis depends strongly on the datasets used for training, validation, and testing. Public datasets such as ISIC, HAM10000, PH2, Derm7pt, BCN20000, PAD-UFES-20, SD-198, Fitzpatrick17k, and DDI have accelerated algorithm development by providing benchmark images for segmentation, classification, multimodal learning, and fairness evaluation. However, these datasets differ substantially in imaging modality, sample size, lesion categories, annotation quality, metadata availability, acquisition setting, and skin-tone representation. Therefore, performance values reported

across studies should not be interpreted without considering dataset characteristics [65–69]. Table 3 summarizes representative datasets commonly used in automated skin lesion and skin cancer detection studies and highlights their strengths, applications, and limitations.

Table 3. Representative datasets commonly used in automated skin lesion and skin cancer detection [65,67,70–72].

Dataset	Image Modality	Dataset Size/Classes	Annotation and Clinical Information	Common Use in AI Studies	Strengths	Key Limitations
ISIC Challenge/ISIC Archive	Mainly dermoscopic images	ISIC 2019 includes 25,331 images across nine diagnostic categories	Diagnostic labels; segmentation masks available in selected challenge tasks	Segmentation, classification, benchmarking, challenge-based comparison	Large international benchmark; widely used for model comparison	Mostly dermoscopic; class imbalance; performance may not generalize to clinical or smartphone images; external validation remains necessary (ISIC Challenge https://challenge.isic-archive.com/landing/2019/?utm_source=chatgpt.com)
HAM10000	Dermoscopic images	10,015 dermoscopic images across common pigmented lesion categories	Diagnostic labels; more than half of lesions confirmed by pathology, with others confirmed by follow-up, expert consensus, or confocal microscopy	Multiclass classification, transfer learning, benchmarking	Large, widely used, multi-source dermoscopic dataset	Class imbalance; limited skin-tone diversity reporting; high internal performance may be dataset-specific (PMC https://pmc.ncbi.nlm.nih.gov/articles/PMC6091241/?utm_source=chatgpt.com)
PH2	Dermoscopic images	200 images: 80 common nevi, 80 atypical nevi, and 40 melanomas	Lesion masks and diagnostic information	Segmentation validation and small-scale classification	Expert masks and well-defined lesion categories	Small sample size; limited lesion diversity; unsuitable alone for training large deep-learning models (PMC https://pmc.ncbi.nlm.nih.gov/articles/PMC8046537/?utm_source=chatgpt.com)
Derm7pt	Dermoscopic and clinical lesion images with metadata	1011 lesion cases	Seven-point checklist criteria, diagnosis, and metadata	Multimodal learning, checklist prediction, clinically interpretable AI	Useful for linking image-based AI with clinically meaningful dermoscopic criteria	Moderate size; missing or variable metadata may affect multimodal reproducibility (ResearchGate https://www.researchgate.net/publication/324381190_7-Point_Checklist_and_Skin_Lesion_Classification_using_Multi-Task_Multi-Modal_Neural_Nets?utm_source=chatgpt.com)
BCN20000	Dermoscopic images	Approximately 18,946–19,424 dermoscopic images collected at Hospital Clínic Barcelona	Diagnostic labels; includes challenging lesion presentations	Large-scale classification and ISIC-related benchmarking	Includes difficult real-world cases such as nail/mucosal lesions, large lesions, and hypopigmented lesions	Single-institution origin; dermoscopy-focused; requires external validation across other centers (PMC https://pmc.ncbi.nlm.nih.gov/articles/PMC11183228/?utm_source=chatgpt.com)
PAD-UFES-20	Clinical images, mostly smartphone-acquired	2298 samples across six/seven lesion categories depending on grouping	Clinical images with patient metadata such as age, lesion location, Fitzpatrick skin type, and lesion diameter	Clinical-image classification, smartphone-based diagnosis, multimodal learning	Real-world clinical-image dataset with patient-level metadata	Smaller than ISIC/HAM10000; geographically specific; acquisition-device variability may affect generalization (PMC https://pmc.ncbi.nlm.nih.gov/articles/PMC7479321/?utm_source=chatgpt.com)

Table 3. Cont.

Dataset	Image Modality	Dataset Size/Classes	Annotation and Clinical Information	Common Use in AI Studies	Strengths	Key Limitations
SD-198	Clinical/macroscopic images	6584 images across 198 skin disease categories	Disease-category labels	Broad skin-disease classification and transfer-learning studies	Covers many dermatological conditions beyond melanoma	Not skin-cancer-specific; many categories have limited images; label and class heterogeneity affect benchmarking
MED-NODE	Clinical/non-dermoscopic images	Commonly reported as 170 images: 70 melanoma and 100 nevi	Diagnostic labels	Binary melanoma/nevus classification using clinical images	Useful for non-dermoscopic evaluation	Very small dataset; limited diversity; not sufficient for standalone deep-learning training (GitHub https://github.com/openmedlab/Awesome-Medical-Dataset/blob/main/resources/MED-NODE.md?utm_source=chatgpt.com)
Fitzpatrick17k	Clinical dermatology images	16,577 clinical images with skin condition and Fitzpatrick skin-type labels	Diagnosis labels and Fitzpatrick skin-type annotations	Fairness analysis, skin-tone bias evaluation, clinical-image classification	Important dataset for evaluating skin-type-related performance disparities	Not limited to skin cancer; class imbalance and skin-type imbalance remain concerns (GitHub https://github.com/mattgroh/fitzpatrick17k?utm_source=chatgpt.com)
DDI: Diverse Dermatology Images	Clinical images	656 images from diverse skin tones and uncommon diseases	Expert-curated and pathologically confirmed images	Fairness testing and external validation of dermatology AI	Valuable for evaluating model performance across diverse skin tones	Small dataset; mainly useful for fairness testing and external validation rather than large-scale training (arXiv https://arxiv.org/abs/2203.08807?utm_source=chatgpt.com)

All websites are accessed on 23 June 2019.

4. Computer-Vision Workflow for Automated Skin Lesion Analysis

Automated skin lesion and skin cancer detection generally follows a sequential computer-vision workflow that converts raw dermatological images into clinically interpretable diagnostic outputs. This workflow connects image acquisition, preprocessing, lesion segmentation, feature extraction, classification, explainable AI, validation, and clinical decision support. As shown in Figure 3, the complete pipeline begins with dermoscopic, clinical, or smartphone-based image acquisition and proceeds through several computational stages before generating an output that can support dermatologist-assisted screening, triage, or diagnostic decision-making.

The first stage is image acquisition and data curation. Skin lesion images may be collected using dermatoscopes, clinical cameras, or smartphone devices. The quality of the acquired image strongly influences downstream model performance because dermatological images are sensitive to illumination variation, camera resolution, focus, acquisition angle, skin reflection, hair, ruler markings, and background skin texture. In addition to image collection, data curation includes diagnostic labeling, expert annotation, train-validation-test data splitting, and handling of class imbalance. Robust partitioning is essential because image-level random splitting or duplicate leakage may inflate reported model performance. Therefore, patient-level separation and careful dataset organization are required before model development.

The second phase is preprocessing, the goal of which is to enhance the image and reduce non-diagnostic artifacts. Typical preprocessing techniques include hair removal, normalization of colors, increasing the contrast of the image, scaling, cutting, reducing noise, and eliminating artifacts. Preprocessing is especially vital in skin lesions since the lesion image may differ significantly depending on the imaging equipment, lighting

conditions, location of the lesion, and color of the skin. Nevertheless, preprocessing should be done with caution and transparency because it could distort the clinical structures of the lesion.

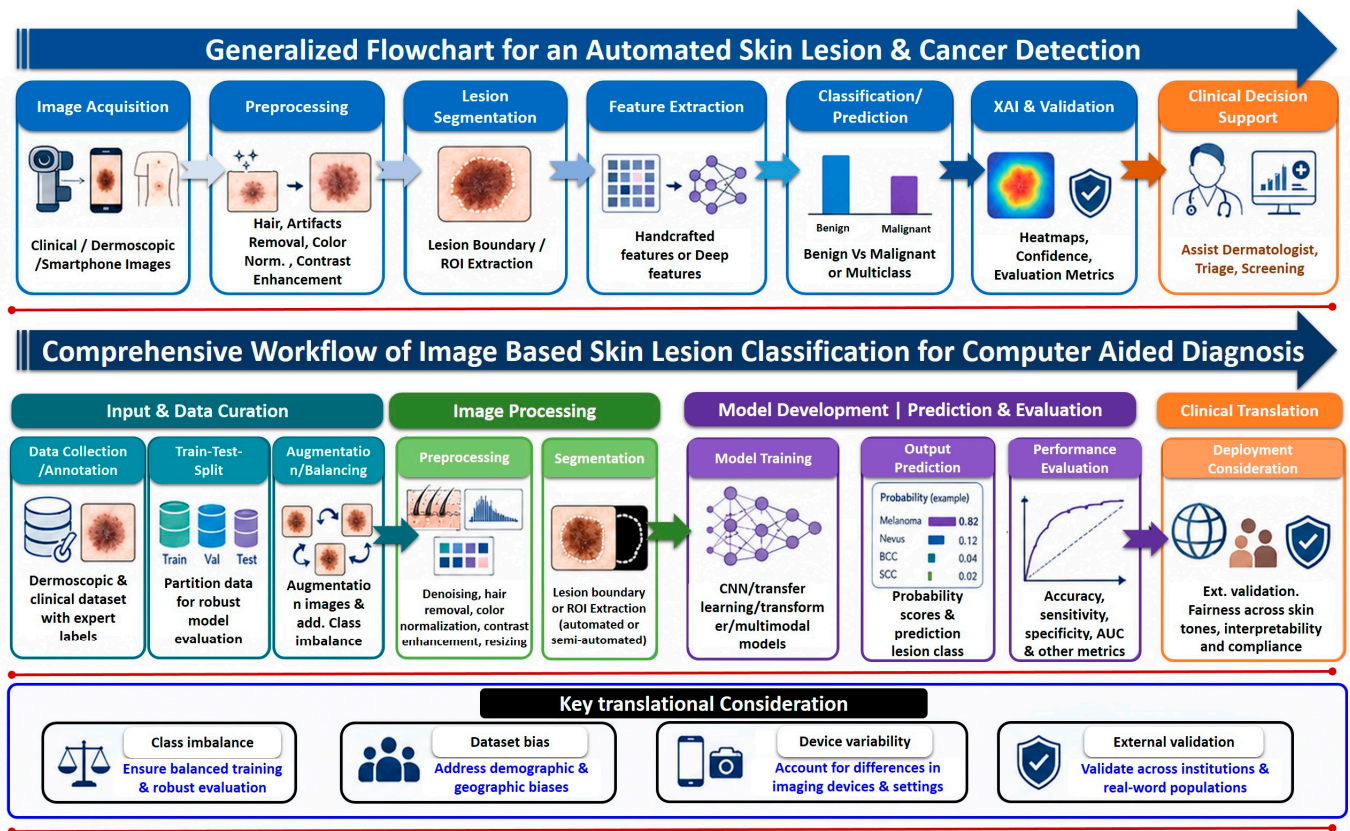


Figure 3. (Top) A Generalized flowchart for an automated skin lesion and cancer detection system and (bottom) a comprehensive workflow of image-based skin lesion classification for computer-aided diagnosis.

The next step involves lesion segmentation, where the boundaries of the lesion or the region of interest are defined. Segmentation helps to extract the lesion area from the rest of the normal skin around it for easier feature extraction and classification. Conventional approaches for segmentation involve thresholding, region growing, active contours, and morphological techniques, while modern approaches make use of U-Net, attention networks, residual networks, and a hybrid CNN–transformer architecture. Segmentation becomes critical when dealing with highly irregular lesions, lesions with low contrast and lesions with partial occlusion. Nevertheless, segmentation is heavily dependent on the quality of annotation, lesion boundaries, and expert validation masks.

Following segmentation, the process continues with feature extraction and modeling. Previous computer-aided diagnosis systems utilized features extracted manually in accordance with the asymmetry, borders, color, texture, and shape characteristics of lesions. Modern artificial intelligence systems rely on a deep neural network architecture for the automatic learning of hierarchical image features from lesion images. The use of CNNs, transfer learning, ensembles, attention mechanisms, and vision transformers is common at this stage. In more advanced AI systems, not only image-based features but also clinical metadata, such as age, gender, body part location, lesion history, and risk factors, can be incorporated into the diagnosis process [73–76].

Classification or prediction is the next phase. The type of classification done will depend on the purpose of the research. The model may either perform binary classification such as benign-versus-malignant prediction or multiclass classification involving

melanoma, nevi, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, dermo-fibroma, and other types of lesions. The output of models typically includes probabilities and diagnosis classes. Various measures used to evaluate the performance of the model include AUC, accuracy, sensitivity, specificity, precision, F1-measure, the Dice similarity coefficient, and Jaccard index. However, these metrics must be interpreted in relation to dataset composition, class imbalance, image modality, split strategy, skin-tone representation, and the external validation setting [46].

The last computational step concerns explainable AI and validation. The use of methods like Grad-CAM, saliency maps, attention maps, heatmaps, and confidence scoring can aid in understanding which parts of the image contributed to the model output. Such methods could increase physicians' trust in models but cannot serve as proof of reliability in a clinical setting without validation with expert judgment and consideration of the lesion region. Validation should encompass not only internal testing but also cross-dataset testing and multi-institutional testing. External validation is especially crucial as a system that performs well on dermoscopic datasets might have different performance characteristics when tested on smartphone images, dark skin types, or rare lesion variants [38,77–79].

The clinical end point of this workflow is decision support. AI-driven solutions should help dermatologists, general practitioners, and tele-dermatology platforms in terms of screening, triaging, prioritizing referrals, lesion tracking, and risk stratification. They should not replace clinical expertise completely. An effective solution should not only provide the diagnosis but also estimate the level of confidence of the prediction, estimate uncertainty, generate an explanation map, and suggest referring a case to an expert in a situation of uncertainty and high-risk predictions. This is why the presented workflow in Figure 3 reveals the necessity of integrating technical and clinical aspects of AI model development and deployment.

5. AI-Based Methodological Approaches for Automated Skin Lesion Analysis

From the computer vision workflow discussed in Section 4, we can deduce that skin lesion analysis using automation does not constitute an isolated step but rather a chain that consists of image capture, image preprocessing, skin lesion localization, feature extraction, classification, explanation, validation, and clinical decision-making. In this workflow, segmentation, classification, generation, and multimodal approaches have complementary functions. Segmentation localizes the lesion area and minimizes the effect of the normal skin around the lesions or any image artifacts. Classification maps the image features into diagnostic decisions such as benign or malignant lesions. Generative approaches handle the problems of unbalanced classes and limited data by generating synthetic or augmented skin lesion images, while multimodal approaches integrate image-based and clinical metadata information.

Consequently, the sections that follow present the different methodologies in automated skin lesion and cancer detection. Rather than viewing the Dice, Jaccard, AUC, accuracy, sensitivity, and specificity scores provided by researchers as comparable across different studies, the following discussion explains these measures in the context of the dataset used, validation procedure, external evaluation, variance/uncertainty reporting, and dataset heterogeneity. The organization of this discussion bridges the workflow explained in Section 4 and the improved comparison tables presented below.

5.1. Segmentation Methods

Lesion segmentation forms the basis of automatic skin lesion analysis since it defines the region of interest that contains the features used in diagnostics. Accurate segmenta-

tion enables the distinction between the lesion and normal skin tissue, hair, ruler lines, illumination variations, and other background artifacts. This becomes especially crucial for cases where the lesion is irregular, has low contrast, is heterogeneous, or is partially hidden. For conventional computer-assisted diagnostic systems, lesion segmentation was usually accomplished through techniques such as thresholding, regional growth, active contours, edge detection, and morphology. Despite being computationally easy, they were very prone to variations due to illumination conditions, lesion color variations, presence of artifacts, and manually set parameters.

The problem of detecting lesion boundaries using deep-learning-based segmentation techniques was significantly enhanced. Encoder–decoder-based networks, such as the U-Net architecture, along with its variations, gained popularity due to their ability to incorporate the task of feature extraction as well as spatial localization by making use of skip connections. Attention mechanism, residual layers, feature aggregation at multiple scales, adversarial training, and hybrid CNN–transformers were some of the more recent improvements for better localization of the boundary with a broader context [80–85].

However, segmentation performance is strongly affected by annotation quality and validation design. Dice and Jaccard scores depend not only on model quality but also on how lesion masks are generated. A polygonal expert mask, a rough manual outline, and a consensus annotation may produce different performance estimates. Similarly, internal testing on ISIC or PH2 does not necessarily indicate that the model will generalize to clinical or smartphone images. For this reason, segmentation results should be interpreted together with dataset type, annotation protocol, cross-dataset validation, and external testing status. A traceability-enhanced comparison of representative segmentation studies is presented in Table 4, which reports the method, dataset, validation design, extracted metrics, availability of confidence/variance measures, external-validation status, and dataset-level limitations.

Table 4. Comparison of representative skin lesion segmentation studies.

Study/Link	Method	Dataset/Image Type	Validation Design	Reported Metrics	CI/SD/ Variance	External Validation	Dataset-Level Heterogeneity/Limitation
Al-Masni et al., 2018 [86]	Full-resolution convolutional network	ISBI 2017, PH2; dermoscopic images	Internal testing on benchmark datasets	ISBI 2017: Jaccard 77.11%, accuracy 94.03%; PH2: Jaccard 84.79%, accuracy 95.08%	NR	Partial cross-dataset reporting using PH2	Dataset size and annotation protocols differ between ISBI and PH2; limited demographic metadata
Vesal et al., 2018 [87]	SkinNet, modified U-Net	ISBI 2017; dermoscopic images	5-fold cross-validation/challenge test setting	Dice 85.10%, Jaccard 76.67%, sensitivity 93.0%	NR	No independent clinical cohort	Good benchmark performance, but limited evidence for device/population generalization
Goyal et al., 2019 [88]	Ensemble Mask R-CNN + DeepLabv3+	ISIC 2017; dermoscopic images	ISIC 2017 train/test setting	Jaccard 79.58%; class-wise accuracy: benign 95.60%, melanoma 90.78%, seborrheic keratosis 91.29%	NR	No	Useful class-level reporting, but no confidence intervals or prospective testing
Sarker et al., 2021 [89]	SLSNet lightweight GAN-based segmentation	ISBI 2017 and ISIC 2018; dermoscopic images	Benchmark-based testing	Accuracy 97.61%, Dice 90.63%, Jaccard 81.98%; >110 FPS on GTX1080Ti	NR	No independent clinical validation	Strong efficiency reporting, but external generalization and fairness not assessed
Innani et al., 2023 [90]	Efficient-GAN/MGAN lesion segmentation	ISIC 2018; dermoscopic images	Internal benchmark evaluation	Dice 90.1%, Jaccard 83.6%, accuracy 94.5%	NR	No	Strong internal segmentation performance; limited evidence under external domain shift

Overall, segmentation methods have evolved from rule-based boundary detection to deep learning and attention-guided architectures. Nevertheless, their clinical reliability remains limited by inconsistent annotation standards, lack of external validation, and limited testing on diverse skin tones and imaging devices. Future segmentation studies should report mask-generation protocols, inter-annotator variability, patient-level data splitting, confidence intervals, and cross-dataset performance.

5.2. Classification Methods

Classification is the central diagnostic stage of automated skin lesion analysis. After image preprocessing and, in many pipelines, lesion segmentation, classification models assign diagnostic labels or risk scores to the lesion. The task may be binary, such as benign versus malignant classification, or multiclass, involving melanoma, nevus, basal cell carcinoma, squamous cell carcinoma, actinic keratosis, seborrheic keratosis, dermatofibroma, vascular lesions, and other lesion categories. Classification models are clinically important because their outputs can support screening, triage, referral prioritization, and dermatologist decision support.

In previous systems, features derived from dermatological guidelines, such as asymmetry, border irregularity, color variance, diameter, texture, and shape attributes, were manually crafted. These features were typically coupled with classic machine learning algorithms like support vector machine, K-nearest neighbor, decision tree, random forest, and artificial neural network. Though valuable in the past, manual feature selection techniques had poor performance in the presence of variability in illumination, imaging devices, skin color, lesion appearance, and segmentation [91–93].

The adoption of CNN revolutionized classification from manually designed feature extraction techniques to feature extraction through end-to-end learning. The following CNN-based models were used to perform classification tasks on several image datasets such as ISIC, HAM10000; that is, the Inception model, ResNet, DenseNet, VGG, EfficientNet, and MobileNet, among others. Transfer learning has gained popularity due to the unavailability of annotations for medical images compared with natural images. Transformer models and CNN–transformer hybrid models have also been tested for classification due to their ability to capture long-range dependency and global information in the image [48,64].

Even with good AUC and accuracy, classification results need to be carefully considered. Many studies apply internal validation, and even random splits by images may lead to leakage of information. The problem of class imbalance is also significant because of an excessive number of benign lesions and the low frequency of melanoma and other rare skin cancer types. In this way, accuracy could be very high despite low sensitivity for malignant classes. Thus, in addition to accuracy and AUC, classification studies have to include data about sensitivity, specificity, confidence intervals, metrics per class, calibration, external validation, and analysis of subgroups according to the Fitzpatrick skin type. Table 5 presents selected examples of classification studies with a focus on the model architecture, dataset, validation method, metric traceability, comparison of statistics, and limitations of datasets.

Table 5. Comparison of representative skin lesion classification studies.

Study/Link	Model/Architecture	Dataset/Image Type	Validation Design	Reported Metrics	CI/SD/Variance	External Validation	Statistical/Comparative Analysis	Dataset-Level Heterogeneity/Limitation
Esteva et al., 2017 [92]	Deep CNN trained end-to-end	129,450 clinical images; 2032 diseases	Tested against 21 board-certified dermatologists on biopsy-proven images	Performance comparable to dermatologists for keratinocyte carcinoma and melanoma classification	NR in accessible abstract-level reporting	No broad multi-institutional prospective validation	Human-expert comparison included	Large training set, but demographic and device-level heterogeneity incompletely reported

Table 5. Cont.

Study/Link	Model/Architecture	Dataset/Image Type	Validation Design	Reported Metrics	CI/SD/Variance	External Validation	Statistical/Comparative Analysis	Dataset-Level Heterogeneity/Limitation
Gouda et al., 2022 [94]	CNN, ResNet50, InceptionV3, Inception-ResNet	ISIC 2018; dermoscopic images	Internal validation	CNN accuracy 83.2%; ResNet50 83.7%; InceptionV3 85.8%; Inception-ResNet 84.0%	NR	No	Architecture comparison included	Internal benchmark performance; limited fairness and external validation
Shetty et al., 2022 [95]	Machine-learning models and CNN	HAM10000; dermoscopic images	10-fold cross-validation	Highest CNN accuracy 95.18%	NR	No	CNN compared with machine-learning classifiers	HAM10000 class imbalance noted; no independent external cohort
Khan and Khan, 2023 [96]	SkinViT transformer model	Three melanoma/non-melanoma datasets	Internal testing across three datasets	Accuracy: 0.9109, 0.8611, 0.8911; AUC: 0.9711, 0.9459, 0.9595	NR	No true external clinical validation	Compared with Efficient-NetV2, MaxViT, MobileViTV2, and ViT	Dataset-level differences exist, but prospective validation and skin-tone subgroup analysis not reported
Tschandl et al., 2020 [97]	CNN-based AI support for human-computer collaboration	Dermoscopic images/telemedical setting	Reader study with AI support	Reported improvement in human-AI decision-making; statistical testing used	Statistical testing reported; detailed CI should be extracted from full text	Clinical-reader setting but not full prospective deployment	Two-sided paired t-tests with Holm-Bonferroni correction reported	Strong clinical relevance, but AI performance depends on reader interaction and setting

5.3. Generative Models and Multimodal Approaches

Generative modeling and multimodal learning are two of the essential improvements made to the segment/classify workflow. They help solve two major problems in skin lesion AI detection: the shortage of annotated data and lack of clinical context. The publicly available skin datasets are not balanced enough since there is limited content on malignant lesions, few rare cases, and low coverage of dark skin. Generative models try to improve the data balance by generating additional images or improving masks for lesions. Multimodal models are trying to make the AI systems more clinically realistic by integrating the images of lesions together with patient data [98,99].

Generative adversarial networks have been applied in the synthesis of dermoscopy images, in the refinement of segmentation masks, in balancing class weights and in generating underrepresented classes of lesions. In conditional GANs, one can generate images using class labels or segmentation masks, whereas some of the other GAN applications target high-resolution generation of lesion images or improvement of the segmentation process. While such applications might help improve the performance of classification and segmentation within the dataset, the generated synthetic images might contain unrealistic textures or repeating patterns that are not clinically meaningful. Learning such synthetic features can lead the classifier to perform very well internally but poorly during external validation. Hence, such models need to be evaluated by their performance, as well as by external validation, fidelity, diversity, and dermatologist evaluation metrics [90,100–102].

Another relevant area is multimodal learning, since dermatologists do not make diagnoses based on image data only. Clinical diagnosis includes age, localization, gender, risks, symptoms, progression of lesions, and dermoscopy criteria. Multimodal AI systems combine image features with metadata using early fusion, late fusion, attention-based fusion, or transformer-based fusion. Studies using datasets such as Derm7pt and PAD-UFES-20 suggest that metadata can improve performance compared with image-only models, especially when lesion images alone are ambiguous. However, multimodal learning introduces new challenges, including missing metadata, inconsistent clinical variables, demographic bias, and reduced reproducibility across datasets [71].

Generative and multimodal studies should therefore be interpreted in relation to their validation design and evidence strength. A small improvement in internal AUC after GAN augmentation does not necessarily prove clinical usefulness. Similarly, a multimodal model may perform well on one dataset but fail when metadata fields are missing or collected differently in another institution. Table 6 summarizes representative generative augmentation and multimodal studies, including their approach, dataset or modality, reported impact, availability of statistical information, key heterogeneity concerns, and evidence interpretation.

Table 6. Traceability-enhanced summary of generative augmentation and multimodal studies.

Study/Paper Link	Approach	Dataset/Modality	Reported Performance or Impact	CI/SD/ Statistical Comparison	Key Heterogeneity/Limitation	Evidence Interpretation
Ding et al., 2020/2021 [103]	Conditional GAN for high-resolution dermoscopy synthesis	Dermoscopy images with segmentation/category conditioning	Image synthesis study; not a direct diagnostic AUC/Dice comparison	NR	Synthetic-image fidelity and downstream clinical benefit need separate validation	Useful for augmentation research, but not sufficient as diagnostic evidence
Kawahara et al., 2019 [104]	Multitask multimodal neural network	Derm7pt; clinical + dermoscopic images + metadata; 1011 lesion cases	Comprehensive diagnostic and seven-point-checklist outputs reported	CI NR in accessible summary	Moderate dataset size; missing metadata can affect reproducibility	Important multimodal reference; metrics should be interpreted task-wise
Lyakhov et al., 2023 [105]	Multimodal neural network with modified cross-entropy loss for imbalance	Heterogeneous dermatological data	Accuracy 85.19%; McNemar statistical significance reported as $p = 0.001$ in comparative tests	Statistical comparison reported	Dataset imbalance addressed, but broader external fairness validation remains needed	Stronger than purely narrative multimodal claims because statistical testing is included
Zhang et al., 2023 [106]	Throughout Fusion Transformer/TFormer	Multimodal skin-lesion diagnosis using dermoscopic/clinical images and metadata	Performance reported in original paper; use source-specific metrics only after extraction	CI NR in accessible summary	Multimodal fusion architecture; computational cost and metadata completeness are key concerns	Useful emerging transformer-based multimodal method; requires careful metric traceability

6. Challenges, Limitations, and Future Directions

Although automated skin lesion and skin cancer detection systems have achieved considerable progress during the last decade, several methodological, clinical, and deployment-related limitations continue to restrict their reliable translation into real-world dermatology practice. Many studies report high Dice, Jaccard, AUC, sensitivity, specificity, and accuracy values on curated public datasets; however, these values are often obtained under internal validation settings and may not fully reflect performance across different populations, institutions, imaging devices, skin tones, lesion sub-types, and clinical workflows. Therefore, future development should shift from isolated benchmark optimization toward clinically robust, externally validated, fair, interpretable, and workflow-compatible AI systems.

6.1. Dataset Bias and Benchmark Design

Among some of the main drawbacks of current research on skin lesions using AI techniques is the use of a small number of open-source datasets like ISIC, HAM10000, PH2, Derm7pt, and SD-198. While these have helped push forward the development of computer-based lesion detection methods, there are some constraints associated with their use. This includes factors like datasets having class imbalances, the presence of too many benign lesions, few cases of rare types of cancerous lesions, variable quality of labeling, and a lack of complete clinical data. Moreover, dark skin colors have been underrepresented in these datasets [107].

Future benchmark datasets need to have more diversity from both clinical and demographic perspectives. This will involve incorporating details such as lesion classification, histopathologic confirmation, lesion location, age, gender, scanning modality, setting, and Fitzpatrick skin type. Future benchmark protocols should involve patient-based data splitting, exclusion of duplicate images, fixed training, validation, and test sets, as well as a clear preprocessing pipeline. Reporting the results of future benchmarks needs to move away from aggregated area under the curve and Dice score reports and embrace individual subgroup analysis, calibration, and confidence interval reporting [23,108–110].

6.2. External and Multi-Institutional Validation

The first important issue contributing to the overestimation of performance in studies on the detection of skin lesions using AI is that, most often, researchers apply an internal data split based on the same dataset. The use of the same source of data for both training and testing can lead to a model that performs well only with images taken in the same setting, using the same device and with the same patients.

Future work must emphasize external validation with datasets from other institutions and geographic locations. Cross-dataset validation must become mandatory in order for claims of generalization to be made. Multi-institution validation would help better assess the performance of models in handling differences in image resolution, illumination, appearance of the lesion, camera use, dermoscopy technique, and patient characteristics. External validation in a prospective clinical setting would help ascertain if the use of artificial intelligence would help improve diagnostic decision-making, referral prioritization, screening efficiency, and patient outcomes [111–114].

6.3. Fairness Across Fitzpatrick Skin Types

The diversity of skin tones is one of the main limitations of existing studies on dermatology. Many freely accessible databases have a great proportion of images related to lighter skin tones, especially Fitzpatrick skin types I–III. Therefore, models based on such databases might have a lower efficiency in terms of darker skin tones, leading to potential delays in the detection of lesions among underrepresented groups. This limitation is vital due to the difference in the manifestation of malignant lesions on different skin tones and anatomical locations.

Performance of the model across all Fitzpatrick skin types I–VI should be reported by future researchers if metadata is available. Fairness assessment should include sensitivity, specificity, false-positive and false-negative rates, AUC, and calibration errors for specific subgroups. Datasets should be created to include darker skin tones, rare lesion appearances, acral and mucosal lesions, and anatomically diverse image samples. Additionally, the development of models should consider the effect of domain adaptation, reweighing, balanced sampling, synthetic augmentation, and fairness-aware learning [23,72,113,115].

6.4. Interpretability and Explainable AI

Explainable AI has become increasingly important in automated skin lesion diagnosis because clinicians need to understand whether a model is focusing on clinically meaningful lesion structures or on irrelevant image artifacts. Commonly used explanation methods include Grad-CAM, saliency maps, Integrated Gradients, occlusion sensitivity, attention visualization, and perturbation-based approaches. These methods generate visual maps that highlight image regions contributing to the model prediction. In skin lesion analysis, reliable explanations should ideally correspond to lesion borders, pigment networks, asymmetry, color variegation, vascular structures, ulceration, or other clinically meaningful dermoscopic patterns [116].

However, the reliability of explainable AI methods remains limited. Saliency maps are often noisy and may change substantially with small image perturbations. Grad-CAM can provide more visually interpretable heatmaps but may highlight broad regions rather than precise lesion structures. Integrated Gradients offers pixel-level attribution but depends on baseline-image selection and may be difficult for clinicians to interpret directly. Attention maps from transformer-based models are sometimes interpreted as explanations, but attention weights do not always correspond to causal diagnostic reasoning. Therefore, explanation outputs should not be treated as proof that a model is clinically reliable [38,117].

One key issue is that AI models might learn shortcuts from features not related to lesions, like rulers, marks of ink, hair, edges, color calibration markers, acquisition artifacts, or background skin structure. This could result in high internal accuracy, yet decrease external generalization. Visual explanations may look very convincing even in the case where there are spurious correlations. It means that visual analysis of heatmaps is not enough. Evaluation of explanation quality should be done quantitatively based on the degree of localization overlap with expert lesion masks, deletion–insertion tests, perturbation robustness, pointing game performance, clinically meaningful region detection, and diagnostic structure alignment.

From a clinician’s point of view, explanations should increase trustworthiness, help error detection, and aid decision-making, but not generate nice-looking heatmaps. Clinicians need explanations that are stable, interpretable, and dermatologically motivated. Therefore, future studies should involve not only technical metrics but also clinical evaluations of explainable AI. Future research should consider whether explanations help clinicians detect model errors, increase their diagnostic confidence, avoid false reassurance, and make decisions about referrals. Overall, explainable AI is a necessary but not sufficient aspect of clinical translation [118,119].

6.5. Uncertainty Estimation and Risk-Aware Decision Support

However, most modern AI solutions provide either a class label or a probabilistic output, without giving information about the reliability of that prediction. This is an issue in the medical field because AI might give very reliable-looking outputs based on low-quality images, uncommon types of lesions, out-of-distribution instances, or images from new types of devices. A clinically valuable solution must be able to recognize the cases where it is uncertain and advise on getting an expert opinion instead of providing a possibly unreliable prediction.

Future medical AI solutions need to include uncertainty assessment and risk-based decision support. Methods like MC-dropout, deep ensembles, Bayesian neural networks, temperature scaling, conformal prediction, and out-of-distribution detection can aid in assessing the confidence of a prediction. Rather than simply giving a binary output as to whether a lesion is benign or malignant, the model needs to provide risk values, uncertainty estimates, and recommendations on whether the case should be referred to a dermatologist [120–122].

6.6. Deployment Constraints and Clinical Workflow Integration

The other major limitation is the fact that most of the studies in skin-lesion AI are conducted offline through retrospective datasets. However, deployment in practice requires more than benchmark accuracies achieve. For successful clinical integration of the system, issues such as inference speed, hardware requirements, image quality management, data privacy, user interface design, clinician adoption, medicolegal liabilities, regulation, and integration into EHR or Teledermatology need to be addressed.

Future studies need to focus on evaluating the AI system within the context of the clinical workflow. It needs to be clear that the use of AI will be to support decision-making and not to replace dermatologists. Evaluation of models would involve how well AI can help dermatologists, general practitioners, and Teledermatologists in patient triaging, referral prioritization, monitoring of lesions, and second opinions. Evaluation of the deployment should consider not just the diagnostic accuracy but also issues related to workflow optimization, efficiency, timeliness, clinician confidence, risk of false reassurance, patient safety, and usability [123–125]. Table 7 summarizes the major limitations and corresponding future recommendations for developing clinically reliable skin-lesion AI systems.

Table 7. Key limitations and future recommendations for skin-lesion AI systems.

Limitation	Current Challenge	Future Recommendation
Dataset bias	Public datasets often show class imbalance, limited rare lesions, and under-representation of darker skin tones	Build multi-source datasets with balanced lesion classes, Fitzpatrick skin type, anatomical site, device, and demographic metadata
Weak benchmark design	Random internal splits may inflate performance	Use patient-level splitting, duplicate removal, fixed test sets, and transparent preprocessing protocols
Limited external validation	Many models are evaluated only on internal test sets	Require cross-dataset, multi-institutional, and prospective validation before clinical claims
Skin-tone fairness	Performance is rarely reported across Fitzpatrick skin types	Report AUC, sensitivity, specificity, false-negative rate, and calibration across Fitzpatrick I–VI
Metric overinterpretation	High AUC, Dice, or accuracy may not indicate clinical reliability	Report confidence intervals, threshold-selection strategy, calibration, and subgroup metrics
Limited interpretability	Heatmaps and saliency maps may highlight artifacts or non-lesion regions	Validate explanations against expert lesion regions and clinically meaningful dermoscopic features
Lack of uncertainty estimation	Models may provide confident outputs for poor-quality or unfamiliar images	Add uncertainty scoring, out-of-distribution detection, and dermatologist-referral triggers
Deployment gap	Most systems are tested offline rather than in real clinical workflows	Conduct workflow studies measuring diagnostic support, clinician trust, usability, safety, and patient outcomes
Privacy and scalability	Centralized data collection may be difficult across hospitals	Explore federated learning, privacy-preserving validation, and secure multi-center model development
Resource constraints	Large models may be unsuitable for mobile or low-resource settings	Use lightweight architectures, compression, quantization, and edge-compatible deployment

7. Conclusions

The present review offers a comprehensive, critical analysis of studies of automatic skin lesion and cancer recognition through computer vision and artificial intelligence methods. Overall, deep learning demonstrated great promise for lesion segmentation, image classification, generative data augmentation, and multimodal decision support. Nevertheless, highly cited Dice, Jaccard, AUC, sensitivity, and specificity measures need to be understood carefully due to the presence of internal validity assessment, imbalanced datasets, underrepresentation of skin tones, and lack of confidence intervals reported and external validation in many studies. Through the combination of methodological comparison, risk-of-bias assessment, heterogeneity of datasets, explainability, fairness, and deployability, this review demonstrates that beyond benchmarking accuracy, there is much

more to clinical translation. Future progress in this field will require multi-institutional, heterogeneous datasets, patient-level validation, uncertainty-based approaches, and clinically viable AI systems.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bioengineering13080872/s1>.

Author Contributions: B.S.S., U.R. and D.P.T.: Writing—original draft, performing algorithm design, software, data acquisition, implementation, and analysis of results. S.G.N., M.B.K., P.R. and M.B.: Algorithm design and implementation, analysis of results, writing and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The review data supporting this systematic review are provided as Supplementary Materials.

Acknowledgments: The authors would like to express their sincere gratitude to the Department of Computer Science, Kalinga University, Raipur, and the Department of Electronics and Telecommunication Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, for providing the necessary research infrastructure, laboratory facilities, and continuous academic support during the course of this work. The authors also acknowledge the valuable guidance and encouragement received from their mentors and colleagues, which greatly contributed to the successful completion of this study. In addition, the availability of open-access datasets such as ISIC and PH², and various research tools used in this work, has been instrumental in advancing the analysis and validation of the proposed methodologies. The authors are also thankful to the editorial and review committee members for their time and constructive feedback, which helped in improving the quality and clarity of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Caraviello, C.; Nazzaro, G.; Tavoletti, G.; Boggio, F.; Denaro, N.; Murgia, G.; Passoni, E.; Benzecry Mancin, V.; Marzano, A.V. Melanoma Skin Cancer: A Comprehensive Review of Current Knowledge. *Cancers* **2025**, *17*, 2920. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Trager, M.H.; Geskin, L.J.; Samie, F.H.; Liu, L. Biomarkers in melanoma and non-melanoma skin cancer prevention and risk stratification. *Exp. Dermatol.* **2022**, *31*, 4–12. [\[CrossRef\]](#)
3. Zambrano-Román, M.; Padilla-Gutiérrez, J.R.; Valle, Y.; Muñoz-Valle, J.F.; Valdés-Alvarado, E. Non-Melanoma Skin Cancer: A Genetic Update and Future Perspectives. *Cancers* **2022**, *14*, 2371. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Adegun, A.; Viriri, S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artif. Intell. Rev.* **2021**, *54*, 811–841. [\[CrossRef\]](#)
5. Zhou, L.; Zhong, Y.; Han, L.; Xie, Y.; Wan, M. Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: Insights from the global burden of disease study 1990–2021. *Sci. Rep.* **2025**, *15*, 5996. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Wang, M.; Gao, X.; Zhang, L. Recent global patterns in skin cancer incidence, mortality, and prevalence. *Chin. Med. J.* **2025**, *138*, 185–192. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Dai, Y.; Liu, X.; Tang, H. Global regional and national burden of nonmelanoma skin cancer from 1990 to 2021 based on the global burden of disease 2021 study. *Discov. Oncol.* **2026**, *17*, 570. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Mashoudy, K.D.; Perez, S.M.; Nouri, K. From diagnosis to intervention: A review of telemedicine's role in skin cancer care. *Arch. Dermatol. Res.* **2024**, *316*, 139. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Yang, G.; Luo, S.; Greer, P. Advancements in skin cancer classification: A review of machine learning techniques in clinical image analysis. *Multimed. Tools Appl.* **2025**, *84*, 9837–9864. [\[CrossRef\]](#)
10. Sonune, B.S.; Udaykumar, R.; Nemane, S.G.; Tulaskar, D.P.; Bhaiyya, M.; Haick, H. Interpretable Deep Learning in Dermoscopy: An XAI-Driven Ensemble for Skin Lesion Diagnosis. *IEEE Access* **2026**, *14*, 62126–62142. [\[CrossRef\]](#)
11. Naik, P.P. Cutaneous Malignant Melanoma: A Review of Early Diagnosis and Management. *World J. Oncol.* **2021**, *12*, 7–19. [\[CrossRef\]](#) [\[PubMed\]](#)

12. Benyahia, S.; Meftah, B.; L  zoray, O. Multi-features extraction based on deep learning for skin lesion classification. *Tissue Cell* **2022**, *74*, 101701. [[CrossRef](#)] [[PubMed](#)]
13. Iqbal, I.; Younus, M.; Walayat, K.; Kakar, M.U.; Ma, J. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. *Comput. Med. Imaging Graph.* **2021**, *88*, 101843. [[CrossRef](#)] [[PubMed](#)]
14. Sonune, B.S.; Kumar, R.U.; Sankar, K.; Agrawal, P.S.; Nemane, S.G.; Tulaskar, D.P.; Bhaiyya, M.; Kulkarni, M.B. Interpretable Skin Cancer Identification Using a Hybrid Deep Learning and XAI Framework on HAM10000. *Bioengineering* **2026**, *13*, 677. [[CrossRef](#)] [[PubMed](#)]
15. Ahmad, I.; Amin, J.; IkramUllah Lali, M.; Abbas, F.; Imran Sharif, M. A novel Deeplabv3+ and vision-based transformer model for segmentation and classification of skin lesions. *Biomed. Signal Process. Control* **2024**, *92*, 106084. [[CrossRef](#)]
16. Narendra, M.; Harshini, T.S.; Anbarasi, L.J. Advancing Skin Disease Diagnosis: A Multimodal Approach Utilizing Telegram Api Token Chatbot for Text and Image Analysis in Skin Disease Classification. *IEEE Access* **2024**, *12*, 189009–189023. [[CrossRef](#)]
17. Owida, H.A.; El-Fattah, I.A.; Abuowaida, S.; Alshdaifat, N.; Mashagba, H.A.; Abd Aziz, A.B.; Alzoubi, A.; Larguech, S.; Al-Bawri, S.S. A deep learning-based dual-branch framework for automated skin lesion segmentation and classification via dermoscopic Images. *Sci. Rep.* **2025**, *15*, 37823. [[CrossRef](#)] [[PubMed](#)]
18. Sarhan, A.M.; Ali, H.A.; Yasser, S.; Gobara, M.; Kandil, A.A.; Sherif, G.; Moustafa, E. Achieving high-accuracy skin cancer classification with deep learning optimized by ant colony algorithm. *J. Electr. Syst. Inf. Technol.* **2025**, *12*, 49. [[CrossRef](#)]
19. Noaman, A.; Ahmad, R.; Khan, M.F.; Mohammed, A.S.; Farooq, M.; Adnan, K.M. Beyond binary: Multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis. *Discov. Appl. Sci.* **2024**, *7*, 35. [[CrossRef](#)]
20. Halder, A.; Dalal, A.; Gharami, S.; Wozniak, M.; Ijaz, M.F.; Singh, P.K. A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. *Sci. Rep.* **2025**, *15*, 6268. [[CrossRef](#)] [[PubMed](#)]
21. Ajabani, D.; Shaikh, Z.A.; Yousef, A.; Ali, K.; Albahar, M.A. Enhancing skin lesion classification: A CNN approach with human baseline comparison. *PeerJ Comput. Sci.* **2025**, *11*, e2795. [[CrossRef](#)] [[PubMed](#)]
22. Mehta, D.; Primiero, C.; Betz-Stablein, B.; Nguyen, T.D.; Gal, Y.; Bowling, A.; Haskett, M.; Sashindranath, M.; Bonnington, P.; Mar, V.; et al. Multi-task AI models in dermatology: Overcoming critical clinical translation challenges for enhanced skin lesion diagnosis. *J. Eur. Acad. Dermatol. Venereol.* **2025**, *39*, 2121–2133. [[CrossRef](#)] [[PubMed](#)]
23. Alipour, N.; Burke, T.; Courtney, J. Skin Type Diversity in Skin Lesion Datasets: A Review. *Curr. Dermatol. Rep.* **2024**, *13*, 198–210. [[CrossRef](#)] [[PubMed](#)]
24. Primiero, C.A.; Rezze, G.G.; Caffery, L.J.; Carrera, C.; Podlipnik, S.; Espinosa, N.; Puig, S.; Janda, M.; Soyer, H.P.; Malvehy, J. A Narrative Review: Opportunities and Challenges in Artificial Intelligence Skin Image Analyses Using Total Body Photography. *J. Invest. Dermatol.* **2024**, *144*, 1200–1207. [[CrossRef](#)] [[PubMed](#)]
25. Bhaiyya, M.; Panigrahi, D.; Rewatkar, P.; Haick, H. Role of Machine Learning Assisted Biosensors in Point-of-Care-Testing for Clinical Decisions. *ACS Sens.* **2024**, *9*, 4495–4519. [[CrossRef](#)] [[PubMed](#)]
26. Balakrishnan, P.; Anny Leema, A.; Jothiaruna, N.; Assudani, P.J.; Sankar, K.; Kulkarni, M.B.; Bhaiyya, M. Artificial intelligence for food safety: From predictive models to real-world safeguards. *Trends Food Sci. Technol.* **2025**, *163*, 105153. [[CrossRef](#)]
27. Combalia, M.; Codella, N.; Rotemberg, V.; Carrera, C.; Dusza, S.; Gutman, D.; Helba, B.; Kittler, H.; Kurtansky, N.R.; Liopyris, K.; et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: The 2019 International Skin Imaging Collaboration Grand Challenge. *Lancet Digit. Health* **2022**, *4*, e330–e339. [[CrossRef](#)] [[PubMed](#)]
28. Wen, D.; Soltan, A.; Trucco, E.; Matin, R.N. From data to diagnosis: Skin cancer image datasets for artificial intelligence. *Clin. Exp. Dermatol.* **2024**, *49*, 675–685. [[CrossRef](#)] [[PubMed](#)]
29. Deshmukh, M.T.; Wankhede, P.R.; Chakole, N.; Kale, P.D.; Jadhav, M.R.; Kulkarni, M.B.; Bhaiyya, M. Towards Intelligent Food Safety: Machine Learning Approaches for Aflatoxin Detection and Risk Prediction. *Trends Food Sci. Technol.* **2025**, *161*, 105055. [[CrossRef](#)]
30. Tulaskar, D.P.; Sindhu, B.; Chakole, N.; Parteki, R.; Leema, A.A.; Balakrishnan, P.; Avthanka, A.; Girhe, R.; Kulkarni, M.B.; Bhaiyya, M. AI and ML empowering 5G and shaping the 6G future: Models, metrics, architectures, and applications. *ICT Express* **2026**, *12*, 111–135. [[CrossRef](#)]
31. Daneshjou, R.; Barata, C.; Betz-Stablein, B.; Celebi, M.E.; Codella, N.; Combalia, M.; Guitera, P.; Gutman, D.; Halpern, A.; Helba, B.; et al. Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology: CLEAR Derm Consensus Guidelines from the International Skin Imaging Collaboration Artificial Intelligence Working Group. *JAMA Dermatol.* **2022**, *158*, 90–96. [[CrossRef](#)] [[PubMed](#)]
32. Ricci Lara, M.A.; Rodr  guez Kowalczuk, M.V.; Eliceche, M.L.; Ferraresso, M.G.; Luna, D.R.; Benitez, S.E.; Mazzuocolo, L.D. A dataset of skin lesion images collected in Argentina for the evaluation of AI tools in this population. *Sci. Data* **2023**, *10*, 712. [[CrossRef](#)] [[PubMed](#)]
33. Nemane, S.; Sarad, V.M.; Tulaskar, D.P.; Marode, T.P.; Bhangdiya, V.; Agrawal, L.; Avthankar, A.; Kondylakis, H.; Kulkarni, M.B.; Bhaiyya, M.; et al. Federated Learning in Multimodal Healthcare Diagnostics: Privacy-Preserving AI for Biomedical Imaging, Electronic Health Records, Wearables, and Clinical Decision Support. *Arch. Comput. Methods Eng.* **2026**. [[CrossRef](#)]

34. Wu, Y.; Chen, B.; Zeng, A.; Pan, D.; Wang, R.; Zhao, S. Skin Cancer Classification With Deep Learning: A Systematic Review. *Front. Oncol.* **2022**, *12*, 893972. [[CrossRef](#)] [[PubMed](#)]
35. Haggenmüller, S.; Maron, R.C.; Hekler, A.; Utikal, J.S.; Barata, C.; Barnhill, R.L.; Beltraminelli, H.; Berking, C.; Betz-Stablein, B.; Blum, A.; et al. Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts. *Eur. J. Cancer* **2021**, *156*, 202–216. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, J.; Zhong, F.; He, K.; Ji, M.; Li, S.; Li, C. Recent Advancements and Perspectives in the Diagnosis of Skin Diseases Using Machine Learning and Deep Learning: A Review. *Diagnostics* **2023**, *13*, 3506. [[CrossRef](#)] [[PubMed](#)]
37. Noronha, S.S.; Mehta, M.A.; Garg, D.; Kotecha, K.; Abraham, A. Deep Learning-Based Dermatological Condition Detection: A Systematic Review with Recent Methods, Datasets, Challenges, and Future Directions. *IEEE Access* **2023**, *11*, 140348–140381. [[CrossRef](#)]
38. Hauser, K.; Kurz, A.; Haggenmüller, S.; Maron, R.C.; von Kalle, C.; Utikal, J.S.; Meier, F.; Hobelsberger, S.; Gellrich, F.F.; Sergon, M.; et al. Explainable artificial intelligence in skin cancer recognition: A systematic review. *Eur. J. Cancer* **2022**, *167*, 54–69. [[CrossRef](#)] [[PubMed](#)]
39. Lyakhova, U.A.; Lyakhov, P.A. Systematic review of approaches to detection and classification of skin cancer using artificial intelligence: Development and prospects. *Comput. Biol. Med.* **2024**, *178*, 108742. [[CrossRef](#)] [[PubMed](#)]
40. Miller, I.; Rosic, N.; Stapelberg, M.; Hudson, J.; Coxon, P.; Furness, J.; Walsh, J.; Climstein, M. Performance of Commercial Dermatoscopic Systems That Incorporate Artificial Intelligence for the Identification of Melanoma in General Practice: A Systematic Review. *Cancers* **2024**, *16*, 1443. [[CrossRef](#)] [[PubMed](#)]
41. Naseri, H.; Safaei, A.A. Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: A systematic literature review. *BMC Cancer* **2025**, *25*, 75. [[CrossRef](#)] [[PubMed](#)]
42. Mehta, P.; Shah, B. Review on Techniques and Steps of Computer Aided Skin Cancer Diagnosis. *Procedia Comput. Sci.* **2016**, *85*, 309–316. [[CrossRef](#)]
43. Dildar, M.; Akram, S.; Irfan, M.; Khan, H.U.; Ramzan, M.; Mahmood, A.R.; Alsaiari, S.A.; Saeed, A.H.M.; Alraddadi, M.O.; Mahnashi, M.H. Skin Cancer Detection: A Review Using Deep Learning Techniques. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5479. [[CrossRef](#)] [[PubMed](#)]
44. Höhn, J.; Hekler, A.; Krieghoff-Henning, E.; Kather, J.N.; Utikal, J.S.; Meier, F.; Gellrich, F.F.; Hauschild, A.; French, L.; Schlager, J.G.; et al. Integrating Patient Data into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J. Med. Internet Res.* **2021**, *23*, e20708. [[CrossRef](#)] [[PubMed](#)]
45. Jones, O.T.; Matin, R.N.; van der Schaar, M.; Prathivadi Bhayankaram, K.; Ranmuthu, C.K.I.; Islam, M.S.; Behiyat, D.; Boscott, R.; Calanzani, N.; Emery, J.; et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: A systematic review. *Lancet Digit. Health* **2022**, *4*, e466–e476. [[CrossRef](#)] [[PubMed](#)]
46. Debelee, T.G. Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review. *Diagnostics* **2023**, *13*, 3147. [[CrossRef](#)] [[PubMed](#)]
47. Nazari, S.; Garcia, R. Automatic Skin Cancer Detection Using Clinical Images: A Comprehensive Review. *Life* **2023**, *13*, 2123. [[CrossRef](#)] [[PubMed](#)]
48. Salinas, M.P.; Sepúlveda, J.; Hidalgo, L.; Peirano, D.; Morel, M.; Uribe, P.; Rotemberg, V.; Briones, J.; Mery, D.; Navarrete-Dechent, C. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. *npj Digit. Med.* **2024**, *7*, 125. [[CrossRef](#)] [[PubMed](#)]
49. Ang, X.L.; Oh, C.C. The Use of Artificial Intelligence for Skin Cancer Detection in Asia-A Systematic Review. *Diagnostics* **2025**, *15*, 939. [[CrossRef](#)] [[PubMed](#)]
50. Karimzadhash, S.; Ghodous, S.; Robati, R.M.; Abbaspour, E.; Goldust, M.; Zaresharifi, N.; Zaresharifi, S. Performance of Artificial Intelligence in Skin Cancer Detection: An Umbrella Review of Systematic Reviews and Meta-Analyses. *Int. J. Dermatol.* **2026**, *65*, 69–85. [[CrossRef](#)] [[PubMed](#)]
51. Rubatto, M.; Sciamarrelli, N.; Borriello, S.; Pala, V.; Mastorino, L.; Tonella, L.; Ribero, S.; Quaglini, P. Classic and new strategies for the treatment of advanced melanoma and non-melanoma skin cancer. *Front. Med.* **2023**, *9*, 959289. [[CrossRef](#)] [[PubMed](#)]
52. Fujimura, T.; Aiba, S. Significance of Immunosuppressive Cells as a Target for Immunotherapies in Melanoma and Non-Melanoma Skin Cancers. *Biomolecules* **2020**, *10*, 1087. [[CrossRef](#)] [[PubMed](#)]
53. Longo, C.; Guida, S.; Mirra, M.; Pampena, R.; Ciardo, S.; Bassoli, S.; Casari, A.; Rongioletti, F.; Spadafora, M.; Chester, J.; et al. Dermoscopy and reflectance confocal microscopy for basal cell carcinoma diagnosis and diagnosis prediction score: A prospective and multicenter study on 1005 lesions. *J. Am. Acad. Dermatol.* **2024**, *90*, 994–1001. [[CrossRef](#)] [[PubMed](#)]
54. Roky, A.H.; Islam, M.M.; Ahasan, A.M.F.; Mostaq, M.S.; Mahmud, M.Z.; Amin, M.N.; Mahmud, M.A. Overview of skin cancer types and prevalence rates across continents. *Cancer Pathog. Ther.* **2025**, *3*, 89–100. [[CrossRef](#)] [[PubMed](#)]
55. Bhatt, H.; Shah, V.; Shah, K.; Shah, R.; Shah, M. State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review. *Intell. Med.* **2023**, *3*, 180–190. [[CrossRef](#)]

56. Hwang, J.C.; Peacker, B.L.; Hartman, R.I. Screening and novel diagnostic technologies for melanoma: An update. *Melanoma Manag.* **2025**, *12*, 2536999. [[CrossRef](#)] [[PubMed](#)]
57. Bakkouri, I.; Afdel, K. Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. *Multimed. Tools Appl.* **2020**, *79*, 20483–20518. [[CrossRef](#)]
58. Khattar, S.; Kaur, R. Computer assisted diagnosis of skin cancer: A survey and future recommendations. *Comput. Electr. Eng.* **2022**, *104*, 108431. [[CrossRef](#)]
59. Mahmoud, N.M.; Soliman, A.M. Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. *Sci. Rep.* **2024**, *14*, 9749. [[CrossRef](#)] [[PubMed](#)]
60. Attallah, O. Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. *Comput. Biol. Med.* **2024**, *178*, 108798. [[CrossRef](#)] [[PubMed](#)]
61. Adla, D.; Reddy, G.V.R.; Nayak, P.; Karuna, G. Deep learning-based computer aided diagnosis model for skin cancer detection and classification. *Distrib. Parallel Databases* **2022**, *40*, 717–736. [[CrossRef](#)]
62. Huang, H.-Y.; Hsiao, Y.-P.; Karmakar, R.; Mukundan, A.; Chaudhary, P.; Hsieh, S.-C.; Wang, H.-C. A Review of Recent Advances in Computer-Aided Detection Methods Using Hyperspectral Imaging Engineering to Detect Skin Cancer. *Cancers* **2023**, *15*, 5634. [[CrossRef](#)] [[PubMed](#)]
63. Malibari, A.A.; Alzahrani, J.S.; Eltahir, M.M.; Malik, V.; Obayya, M.; Al Duhayyim, M.; Lira Neto, A.V.; de Albuquerque, V.H.C. Optimal deep neural network-driven computer aided diagnosis model for skin cancer. *Comput. Electr. Eng.* **2022**, *103*, 108318. [[CrossRef](#)]
64. Meedeniya, D.; De Silva, S.; Gamage, L.; Isuranga, U. Skin cancer identification utilizing deep learning: A survey. *IET Image Process.* **2024**, *18*, 3731–3749. [[CrossRef](#)]
65. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [[CrossRef](#)] [[PubMed](#)]
66. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.S.; Rozeira, J. PH2—A dermoscopic image database for research and benchmarking. In *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; IEEE: Piscataway, NJ, USA, 2013; pp. 5437–5440. [[CrossRef](#)] [[PubMed](#)]
67. Hernández-Pérez, C.; Combalia, M.; Podlipnik, S.; Codella, N.C.F.; Rotemberg, V.; Halpern, A.C.; Reiter, O.; Carrera, C.; Barreiro, A.; Helba, B.; et al. BCN20000: Dermoscopic Lesions in the Wild. *Sci. Data* **2024**, *11*, 641. [[CrossRef](#)] [[PubMed](#)]
68. Cockayne, M.J.; Ortolani, M.; Al-Bander, B. DermFormer: Nested multi-modal vision transformers for robust skin cancer detection. *Pattern Anal. Appl.* **2025**, *28*, 194. [[CrossRef](#)]
69. Sun, X.; Yang, J.; Sun, M.; Wang, K. A Benchmark for Automatic Visual Classification of Clinical Skin Disease Images. In *Computer Vision—ECCV 2016; Lecture Notes in Computer Science*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 206–222.
70. Hwang, Y.N.; Seo, M.J.; Kim, S.M. A Segmentation of Melanocytic Skin Lesions in Dermoscopic and Standard Images Using a Hybrid Two-Stage Approach. *BioMed Res. Int.* **2021**, *2021*, 5562801. [[CrossRef](#)] [[PubMed](#)]
71. Pacheco, A.G.C.; Lima, G.R.; Salomão, A.S.; Krohling, B.; Biral, I.P.; de Angelo, G.G.; Alves, F.C.R.J.; Esgario, J.G.M.; Simora, A.C.; Castro, P.B.C.; et al. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data Brief* **2020**, *32*, 106221. [[CrossRef](#)] [[PubMed](#)]
72. Daneshjou, R.; Vodrahalli, K.; Novoa, R.A.; Jenkins, M.; Liang, W.; Rotemberg, V.; Ko, J.; Swetter, S.M.; Bailey, E.E.; Gevaert, O.; et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci. Adv.* **2022**, *8*, eabq6147. [[CrossRef](#)] [[PubMed](#)]
73. Assudani, P.J.; Bhurgy, A.S.; Kollem, S.; Bhurgy, B.S.; Ahmad, M.O.; Kulkarni, M.B.; Bhaiyya, M. Artificial intelligence and machine learning in infectious disease diagnostics: A comprehensive review of applications, challenges, and future directions. *Microchem. J.* **2025**, *218*, 115802. [[CrossRef](#)]
74. Wei, M.L.; Tada, M.; So, A.; Torres, R. Artificial intelligence and skin cancer. *Front. Med.* **2024**, *11*, 1331895. [[CrossRef](#)] [[PubMed](#)]
75. Abhishek, K.; Jain, A.; Hamarneh, G. Investigating the Quality of DermaMNIST and Fitzpatrick17k Dermatological Image Datasets. *Sci. Data* **2025**, *12*, 196. [[CrossRef](#)] [[PubMed](#)]
76. Joseph, S.; Olugbara, O.O. Preprocessing Effects on Performance of Skin Lesion Saliency Segmentation. *Diagnostics* **2022**, *12*, 344. [[CrossRef](#)] [[PubMed](#)]
77. Pathan, S.; Prabhu, K.G.; Siddalingaswamy, P.C. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review. *Biomed. Signal Process. Control* **2018**, *39*, 237–262. [[CrossRef](#)]
78. Agrawal, L.; Agrawal, P.K.; Agrawal, S.S.; Sonune, M.S.; Kadu, R.K.; Kulkarni, M.B.; Bhaiyya, M. AI-driven multimodal retinal imaging for early detection and risk stratification of vascular and neurodegenerative diseases. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2026**. [[CrossRef](#)] [[PubMed](#)]

79. Assudani, P.J.; Balakrishnan, P.; Anny Leema, A.; George, G.; Avthankar, A.; Tiwari, A.; Bhaiyya, M.; Kulkarni, M.B. Biosensing technologies for foodborne pathogen detection and healthcare: Principles, emerging materials, and intelligent platforms. *Microchim. Acta* **2026**, *193*, 231. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Mirikharaji, Z.; Abhishek, K.; Bissoto, A.; Barata, C.; Avila, S.; Valle, E.; Celebi, M.E.; Hamarneh, G. A survey on deep learning for skin lesion segmentation. *Med. Image Anal.* **2023**, *88*, 102863. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Hammimou, A.; Ezzahori, H.; Boudaoud, A.; Aqil, M. From traditional to deep learning methods for skin lesion segmentation: A literature review. *Sci. Afr.* **2025**, *29*, e02783. [\[CrossRef\]](#)
82. Liu, Z.; Hu, J.; Gong, X.; Li, F. Skin lesion segmentation with a multiscale input fusion U-Net incorporating Res2-SE and pyramid dilated convolution. *Sci. Rep.* **2025**, *15*, 7975. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Hu, G.; Zhu, W.; Liao, X.; Li, Q. WA-NET: Enhanced boundary-aware segmentation of skin lesions via frequency-spatial feature fusion and attention-guided edge refinement. *Sci. Rep.* **2025**, *15*, 41598. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Pennisi, A.; Bloisi, D.D.; Suriani, V.; Nardi, D.; Facchiano, A.; Giampetruzzi, A.R. Skin Lesion Area Segmentation Using Attention Squeeze U-Net for Embedded Devices. *J. Digit. Imaging* **2022**, *35*, 1217–1230. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Le, P.T.; Pham, B.-T.; Chang, C.-C.; Hsu, Y.-C.; Tai, T.-C.; Li, Y.-H.; Wang, J.-C. Anti-Aliasing Attention U-net Model for Skin Lesion Segmentation. *Diagnostics* **2023**, *13*, 1460. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Al-Masni, M.A.; Al-Antari, M.A.; Choi, M.-T.; Han, S.-M.; Kim, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **2018**, *162*, 221–231. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Vesal, S.; Ravikumar, N.; Maier, A. SkinNet: A Deep Learning Framework for Skin Lesion Segmentation. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–3. [\[CrossRef\]](#)
88. Goyal, M.; Oakley, A.; Bansal, P.; Dancey, D.; Yap, M.H. Skin Lesion Segmentation in Dermoscopic Images with Ensemble Deep Learning Methods. *IEEE Access* **2020**, *8*, 4171–4181. [\[CrossRef\]](#)
89. Sarker, M.M.K.; Rashwan, H.A.; Akram, F.; Singh, V.K.; Banu, S.F.; Chowdhury, F.U.H.; Choudhury, K.A.; Chambon, S.; Radeva, P.; Puig, D.; et al. SLSNet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Syst. Appl.* **2021**, *183*, 115433. [\[CrossRef\]](#)
90. Innani, S.; Dutande, P.; Baid, U.; Pokuri, V.; Bakas, S.; Talbar, S.; Baheti, B.; Guntuku, S.C. Generative adversarial networks based skin lesion segmentation. *Sci. Rep.* **2023**, *13*, 13467. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Kassem, M.A.; Hosny, K.M.; Damaševičius, R.; Eltoukhy, M.M. Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review. *Diagnostics* **2021**, *11*, 1390. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Jain, S.; Singhanian, U.; Tripathy, B.; Nasr, E.A.; Aboudaif, M.K.; Kamrani, A.K. Deep Learning-Based Transfer Learning for Classification of Skin Cancer. *Sensors* **2021**, *21*, 8142. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Gouda, W.; Sama, N.U.; Al-Waakid, G.; Humayun, M.; Jhanjhi, N.Z. Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare* **2022**, *10*, 1183. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Shetty, B.; Fernandes, R.; Rodrigues, A.P.; Chengoden, R.; Bhattacharya, S.; Lakshmana, K. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Sci. Rep.* **2022**, *12*, 18134. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Khan, S.; Khan, A. SkinViT: A transformer based method for Melanoma and Nonmelanoma classification. *PLoS ONE* **2023**, *18*, e0295151. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Tschandl, P.; Rinner, C.; Apalla, Z.; Argenziano, G.; Codella, N.; Halpern, A.; Janda, M.; Lallas, A.; Longo, C.; Malvehy, J.; et al. Human-computer collaboration for skin cancer recognition. *Nat. Med.* **2020**, *26*, 1229–1234. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Goceri, E. GAN based augmentation using a hybrid loss function for dermoscopy images. *Artif. Intell. Rev.* **2024**, *57*, 234. [\[CrossRef\]](#)
99. Lei, B.; Xia, Z.; Jiang, F.; Jiang, X.; Ge, Z.; Xu, Y.; Qin, J.; Chen, S.; Wang, T.; Wang, S. Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Anal.* **2020**, *64*, 101716. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Qin, Z.; Liu, Z.; Zhu, P.; Xue, Y. A GAN-based image synthesis method for skin lesion classification. *Comput. Methods Programs Biomed.* **2020**, *195*, 105568. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Duan, J.; Xiong, J.; Li, Y.; Ding, W. Deep learning based multimodal biomedical data fusion: An overview and comparative review. *Inf. Fusion* **2024**, *112*, 102536. [\[CrossRef\]](#)
102. Chakkarapani, V.; Poornapushpakala, S.; Suresh, S. Enhancing Skin Cancer Detection with Multimodal Data Integration: A Combined Approach Using Images and Clinical Notes. *SN Comput. Sci.* **2025**, *6*, 72. [\[CrossRef\]](#)
103. Izu-Belloso, R.M.; Ibarrola-Altuna, R.; Rodriguez-Alonso, A. Generative Adversarial Networks in Dermatology: A Narrative Review of Current Applications, Challenges, and Future Perspectives. *Bioengineering* **2025**, *12*, 1113. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Kawahara, J.; Daneshvar, S.; Argenziano, G.; Hamarneh, G. 7-Point Checklist and Skin Lesion Classification using Multi-Task Multi-Modal Neural Nets. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 538–546. [\[CrossRef\]](#) [\[PubMed\]](#)

105. Lyakhov, P.A.; Lyakhova, U.A.; Kalita, D.I. Multimodal Analysis of Unbalanced Dermatological Data for Skin Cancer Recognition. *IEEE Access* **2023**, *11*, 131487–131507. [\[CrossRef\]](#)
106. Zhang, Y.; Xie, F.; Chen, J. TFormer: A throughout fusion transformer for multi-modal skin lesion diagnosis. *Comput. Biol. Med.* **2023**, *157*, 106712. [\[CrossRef\]](#) [\[PubMed\]](#)
107. Cassidy, B.; Kendrick, C.; Brodzicki, A.; Jaworek-Korjakowska, J.; Yap, M.H. Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Med. Image Anal.* **2022**, *75*, 102305. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Wen, D.; Khan, S.M.; Xu, A.J.; Ibrahim, H.; Smith, L.; Caballero, J.; Zepeda, L.; de Blas Perez, C.; Denniston, A.K.; Liu, X.; et al. Characteristics of publicly available skin cancer image datasets: A systematic review. *Lancet Digit. Health* **2022**, *4*, e64–e74. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Li, Y.; Taylor, M.; Chmielinski, K.S.; Halpern, A.C.; Daneshjou, R.; Lester, J.C.; Rotemberg, V. Improving dataset transparency in dermatologic Artificial Intelligence using a dataset nutrition label. *npj Digit. Med.* **2025**, *8*, 641. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Weir, V.R.; Li, Y.; Gillis, M.C.; Kurtansky, N.R.; Salvador, T.; Halpern, A.C.; Nelson, K.C.; Lester, J.C.; Rotemberg, V. Evaluating skin tone scales for dermatologic dataset labeling: A prospective-comparative study. *npj Digit. Med.* **2025**, *8*, 787. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Martínez-Vargas, E.; Mora-Jiménez, J.; Arguedas-Chacón, S.; Hernández-López, J.; Zavaleta-Monestel, E. The Emerging Role of Artificial Intelligence in Dermatology: A Systematic Review of Its Clinical Applications. *Dermato* **2025**, *5*, 9. [\[CrossRef\]](#)
112. Fernandes, T.R.S.; Teles, A.S.; Fernandes, J.R.N.; Lima, L.D.B.; Sousa, D.L.; Soares, R.d.C.; Teixeira, S.S. External Validation of AI Models for Skin Diseases: A Systematic Review. *IEEE Access* **2025**, *13*, 114411–114427. [\[CrossRef\]](#)
113. Tjiu, J.-W.; Lu, C.-F. Equity and Generalizability of Artificial Intelligence for Skin-Lesion Diagnosis Using Clinical, Dermoscopic, and Smartphone Images: A Systematic Review and Meta-Analysis. *Medicina* **2025**, *61*, 2186. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Marchetti, M.A.; Cowen, E.A.; Kurtansky, N.R.; Weber, J.; Dauscher, M.; DeFazio, J.; Deng, L.; Dusza, S.W.; Haliasos, H.; Halpern, A.C.; et al. Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study). *npj Digit. Med.* **2023**, *6*, 127. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Dowie, T. Exploring the Diagnostic Capability of Artificial Intelligence in Dermatology for Darker Skin Tones: A Narrative Review. *Cureus* **2025**, *17*, e94909. [\[CrossRef\]](#) [\[PubMed\]](#)
116. Chanda, T.; Hauser, K.; Hobelsberger, S.; Bucher, T.-C.; Garcia, C.N.; Wies, C.; Kittler, H.; Tschandl, P.; Navarrete-Dechent, C.; Podlipnik, S.; et al. Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. *Nat. Commun.* **2024**, *15*, 524. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Giavina-Bianchi, M.; Vitor, W.G.; Fornasiero de Paiva, V.; Okita, A.L.; Sousa, R.M.; Machado, B. Explainability agreement between dermatologists and five visual explanations techniques in deep neural networks for melanoma AI classification. *Front. Med.* **2023**, *10*, 1241484. [\[CrossRef\]](#) [\[PubMed\]](#)
118. Ennab, M.; Mcheick, H. Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 12. [\[CrossRef\]](#)
119. Chanda, T.; Haggenmueller, S.; Bucher, T.-C.; Holland-Letz, T.; Kittler, H.; Tschandl, P.; Heppt, M.V.; Berking, C.; Utikal, J.S.; Schilling, B.; et al. Dermatologist-like explainable AI enhances melanoma diagnosis accuracy: Eye-tracking study. *Nat. Commun.* **2025**, *16*, 4739. [\[CrossRef\]](#) [\[PubMed\]](#)
120. Young, A.T.; Fernandez, K.; Pfau, J.; Reddy, R.; Cao, N.A.; von Franque, M.Y.; Johal, A.; Wu, B.V.; Wu, R.R.; Chen, J.Y.; et al. Stress testing reveals gaps in clinic readiness of image-based diagnostic artificial intelligence models. *npj Digit. Med.* **2021**, *4*, 10. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Tabarisaadi, P.; Khosravi, A.; Nahavandi, S. Uncertainty-aware skin cancer detection: The element of doubt. *Comput. Biol. Med.* **2022**, *144*, 105357. [\[CrossRef\]](#) [\[PubMed\]](#)
122. Fayyad, J.; Alijani, S.; Najjaran, H. Empirical validation of Conformal Prediction for trustworthy skin lesions classification. *Comput. Methods Programs Biomed.* **2024**, *253*, 108231. [\[CrossRef\]](#) [\[PubMed\]](#)
123. Heinlein, L.; Maron, R.C.; Hekler, A.; Haggenmüller, S.; Wies, C.; Utikal, J.S.; Meier, F.; Hobelsberger, S.; Gellrich, F.F.; Sergon, M.; et al. Prospective multicenter study using artificial intelligence to improve dermoscopic melanoma diagnosis in patient care. *Commun. Med.* **2024**, *4*, 177. [\[CrossRef\]](#) [\[PubMed\]](#)
124. Sangers, T.E.; Wakkee, M.; Moolenburgh, F.J.; Nijsten, T.; Lugtenberg, M. Towards successful implementation of artificial intelligence in skin cancer care: A qualitative study exploring the views of dermatologists and general practitioners. *Arch. Dermatol. Res.* **2023**, *315*, 1187–1195. [\[CrossRef\]](#) [\[PubMed\]](#)
125. Marsden, H.; Kemos, P.; Venzi, M.; Noy, M.; Maheswaran, S.; Francis, N.; Hyde, C.; Mullarkey, D.; Kalsi, D.; Thomas, L. Accuracy of an artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service. *Front. Med.* **2024**, *11*, 1302363. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.