



## Deep Learning in Dermatology: A Systematic Review of Image Analysis, Disease Classification, and Explainable AI

Dr. Satish Sahu<sup>1</sup>, Abhishek Tiwari<sup>2\*</sup>

<sup>1</sup> Department of Commerce and Management, Dr. C. V. Raman University, Bilaspur, Chhattisgarh, India

<sup>2</sup> Department of Information Technology and Computer Science, Dr. C. V. Raman University, Bilaspur, Chhattisgarh, India

\* Email ID: [akt.champa90@gmail.com](mailto:akt.champa90@gmail.com)

### ARTICLE INFO

*Article type:*

Systematic review

*Article history:*

Received: 28-07-2026

Received in revised form: 04-08-2026

Accepted: 04-08-2026

Available online: 05-08-2026

*Keywords:*

Clinical Decision Support

Deep Learning

Dermatology

Explainable Artificial

Intelligence

Skin Lesion Classification

Vision Transformers

### ABSTRACT

Deep learning is changing the way dermatology assessment is done through skin images, yet the evidence remains scattered across image preprocessing, segmentation, classification, explainability, and clinical validation. This systematic review analysed 40 studies published between 2017 and 2026. Google Scholar, IEEE Xplore, SpringerLink, and MDPI returned 52 records; after removing 12 duplicates, 40 unique studies met the inclusion criteria. Reporting quality was assessed with a five-point checklist. The evidence shows a move from manually selected features and classical CNNs to EfficientNet, attention mechanisms, vision transformers, and hybrid CNN-transformer architectures. ISIC, HAM10000, PH2, and PAD-UFES-20 are the main datasets, while Grad-CAM and its variants are gaining momentum for explaining model attention. Performance measures are difficult to compare because of differences in datasets, class divisions, validation techniques, and reported outcomes. Ongoing challenges include uneven class representation, restricted skin-tone variance, high computational requirements, uncertain calibration, weak external validation, and limited testing of the clinical relevance of explanations. Future work should prioritise diverse multicentre datasets, prospective workflow assessment, subgroup disclosure, lightweight deployment, and clinically valuable explanations.

## 1. INTRODUCTION

### 1.1 Clinical Significance and Diagnostic Burden

Dermatological disorders can be quite serious. In some cases, a patient will suffer from life-threatening malignancies. Also, besides physical symptoms, these conditions can impair one's ability to enjoy life and make them feel isolated, and they will need to visit health services frequently [1, 2]. Skin-cancer image-based diagnostic artificial intelligence has drawn special attention because early detection of suspicious skin lesions can change the referral, biopsy, and treatment plans [3, 4, 5].

Confusing appearance between diseases is still a major problem in the field of diagnostics: lesions caused by diverse factors might have the same colors, textures, borders, and distribution, on the other hand, a single disease type could present itself in various ways according to the site and the skin tone [6, 7, 8]. Therefore, a physical exam must also take into consideration contextual and professional judgment elements and in some cases the clinicians may not agree. Thoroughly judged decision-support tools might prove helpful in extending the role of the first-level assessments in areas that have limited access to specialists, but their role should be that of supporting rather than replacing a clinical assessment [9, 10].

## 1.2 Evolution of Computational Methods in Dermatological Diagnosis

Older computer-aided pipelines were essentially image enhancement, lesion delineation, feature engineering and classification divided into separate steps. The system relied on hand-crafted descriptors such as colour, texture, shape and border irregularity, to make its predictions. As a result, the performance varied according to the choice of features and acquisition conditions [11, 12, 13].

Deep learning has lessened the need for human-created features to an extent, allowing models to learn the features themselves from the input (the dermoscopic or clinical images, that is,). Models such as VGG, ResNet, DenseNet and Inception in case of Convolutional neural networks have been used for segmentation and also classification with some research papers being quite influential that showcased excellent results with the model's ability to distinguish different classes under controlled situations [1, 4, 5, 6]. Although these outcomes are promising, one must bear in mind that just because a model has excellent benchmark accuracy, it does not mean that the model will be safe, effective, or perform well if deployed in regular healthcare settings.

Architectural development since then has focused on large compound-scaled EfficientNet models, attention modules, vision transformers, and hybrid CNN, transformer systems [14, 15, 16, 17]. These models try to either increase parameter efficiency very much or they attempt to achieve a combination of local texture clues with a long-range image context.

## 1.3 Public Datasets and Benchmark Repositories

Public data sets have helped dermatological AI research as they make the results reproducible and comparable. For instance, The International Skin Imaging Collaboration (ISIC) Archive offers dermoscopic images, annotations, and challenge tasks that have directed various model developments [18, 19]. Patient oriented data sets that take into account the medical history also explain why a single isolated lesion may not capture the information that is used daily [20].

Commonly cited skin cancer databases are HAM10000, PH2, PAD-UFES-20, and Fitzpatrick17k [21, 22, 23, 24]. The rationale and the environment of data collection vary: some are geared towards close-up images or dermoscopy, others smartphone pictures or skin-tone inclusion. Inspection of these data sets have revealed that they provide incomplete information about the population involved and an uneven representation. That means a potential risk of inaccurate group performance metrics and disparity in error rate [19, 25, 26].

## 1.4 Clinical Translation Gap and Review Objective

The translational problem mainly rests on the fact that a high mean or average performance may mask a general systematic failure for certain under-represented groups. These problems stem from the limited number of dark skin colours, the incomplete reporting of demographic features, and the concentration in certain geographic regions of the image's collections [9, 19, 21, 25].

Another factor to consider during deployment is the actual limitations. Big CNNs and transformer systems could need memory, computational power, and latency that are tough to meet mobile devices or the services that have limited resources [15, 17, 27, 28]. As a result, minimalist design is only of value if the performance is not reduced in any aspect like sensitivity, calibration, or subgroup performance at the expense of efficiency.

The second gap separates benchmarks on the basis of historical data with real-world assessment. Imbalance between the classes, presence of image artefacts, limited range of diseases, images of the same person shown repeatedly, and limited external testing can result in overestimated generalizability [1, 6, 10, 16, 19]. Likewise, explanations that appear convincing may be unstable or unhelpful unless their accuracy and helpfulness are tested.

Against this background, the article reviews four issues: What dermatology-related segmentation tasks, image analysis and deep learning methods most widely used are? How segmentation, classification, and explanation are assessed; What kinds of research shortcomings make it difficult for us to compare and clinically use the results? Which future research directions would lead us to a better implementation of decision systems that we can trust? The summary deals with preprocessing, lesion segmentation, disease classification, model interpretability, benchmarking, and clinical integration challenges.

## 2. RESEARCH METHODOLOGY

### 2.1 Search Strategy and Information Sources

An exhaustive systematic search was done via Google Scholar, IEEE Xplore, SpringerLink, and MDPI in July 2026. Four concept blocks were tailored to the specific language of each source: (1) "skin disease" together with "deep learning," also "image processing;" (2) "skin lesion" or "melanoma" combined with "CNN" or "ResNet"), in addition, with segmentation; (3) ISIC AND "deep learning" AND "skin classification;" (4) "detection of skin disease" AND "deep learning" AND (challenges review). The results were filtered to get only English-language contents related to image-based dermatology assessment and 2017, 2026 time period as we do not want to limit our search too much. Only the dataset papers published prior to that time period were included if they are necessary to describe the benchmark data.

Search results were gathered together and duplicates for title, author, year, and DOI were checked out in detail. The remaining ones were then evaluated for matching with the initial eligibility criteria before the main evidence was the one containing 40 distinct studies that are included for the descriptive, technical, and thematic synthesis. Papers which formed the background such as methodological papers and a paper which described a new method, Grad-CAM paper, clinical validation reports, and similar materials, were not included in the 40-study bibliometric denominator.

### 2.2 Eligibility Criteria

The eligibility criteria were clarified before the last stage of screening so that the evidence set would match the review questions completely. Only the studies that included image-based detection, segmentation, or classification of skin diseases were included. The reports selected had to present the methods and outcomes in such a detail that the data could be extracted.

**Table 1** operationalizes the inclusion rules applied during screening.

**Table 1. Operational inclusion criteria for study selection.**

| Code | Criterion                  | Operational definition                                                                                                              |
|------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| IC1  | Clinical domain relevance  | Detection, segmentation, or classification of a dermatological condition, including malignant, inflammatory, or infectious disease. |
| IC2  | Technical framework        | Use of machine learning or deep learning, including CNN, EfficientNet, attention, or transformer-based methods.                     |
| IC3  | Input data modality        | Analysis of dermoscopy images, clinical photographs, or smartphone-acquired skin images.                                            |
| IC4  | Chronological scope        | Publication dated 2017–2026, except foundational dataset reports cited for context.                                                 |
| IC5  | Publication type           | Peer-reviewed journal article or conference paper with sufficient scholarly and technical detail.                                   |
| IC6  | Language and accessibility | English-language full text available for assessment.                                                                                |
| IC7  | Performance reporting      | At least one quantitative result, such as accuracy, sensitivity, specificity, F1, AUC, Dice, or Jaccard.                            |
| IC8  | Dataset transparency       | A named public dataset or an adequately described custom dataset.                                                                   |

**Table 2** lists the conditions under which a retrieved record was removed from consideration.

**Table 2. Operational exclusion criteria for study selection.**

| Code | Criterion                     | Operational definition                                                                                                         |
|------|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| EC1  | Duplicate records             | A record matching another item by title, DOI, authorship, year, or clearly identical report.                                   |
| EC2  | Non-primary literature        | Editorial, commentary, letter, book review, or other publication without primary or review evidence relevant to the synthesis. |
| EC3  | Grey literature               | Thesis, dissertation, poster, abstract-only record, or unpublished report.                                                     |
| EC4  | Insufficient technical detail | Insufficient description of the dataset, architecture, evaluation, or outcome for structured extraction.                       |
| EC5  | Non-image-based approaches    | Textual, genomic, or tabular analysis with no dermatological image component.                                                  |
| EC6  | Out-of-scope medical imaging  | Imaging outside dermatology, including CT, MRI, or ultrasound applications unrelated to skin disease.                          |
| EC7  | Non-AI approaches             | Rule-based image processing without a machine-learning or deep-learning component.                                             |

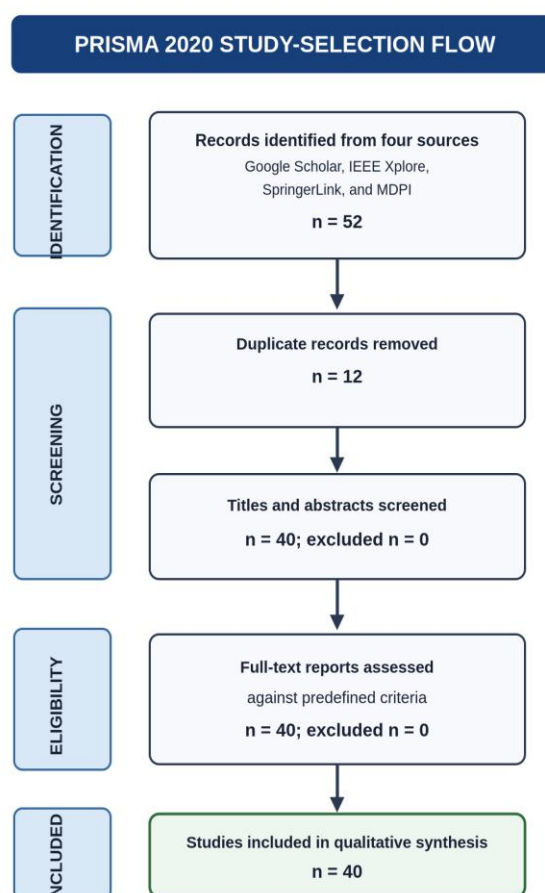
|     |                      |                                   |
|-----|----------------------|-----------------------------------|
| EC8 | Language restriction | Full text unavailable in English. |
|-----|----------------------|-----------------------------------|

### 2.2.1 Rationale for the Eligibility Rules

During the 2017–2026 period, transfer learning, attention mechanisms, transformer models, and explanations techniques were increasingly used in dermatology imaging research. Selecting only those reports which featured identifiable datasets and quantitative results allowed for analyzing inputs, models, evaluation practices, and limitations. Foundations dataset papers outside this time were only kept, when necessary, in order to document the origin or the composition of a benchmark that was a key research tool in the studies covered.

### 2.3 Study Selection Process (PRISMA)

Study selection was described according to PRISMA 2020 [29]. Four different information sources turned up 52 records altogether. After comparing the metadata of the citations, 12 duplicates were discarded, so we had 40 unique records remaining for title- and abstract-screening, and, finally, for full-text assessment.



**Figure 1. PRISMA 2020 flow diagram for identification, screening, eligibility assessment, and inclusion.**

Forty unique records all complied with the eligibility criteria and were included in the qualitative synthesis. At the screening stage, they were grouped in four domains; namely image preprocessing, lesion segmentation, disease classification, and diagnosis or explanation supported by artificial intelligence. Since post deduplication exclusions occurred no more, this is due to the narrow and topic specific search design and is not a proof that the search was complete.

**Figure 1** shows the full numerical audit trail. Besides retained publications, the data covered various diseases, datasets, imaging modalities, model families, and evaluation measures, which made possible the structured comparative analysis while highlighting considerable methodological diversity.

## 2.4 Data Extraction, Quality Appraisal, and Synthesis

A structured extraction form captured bibliographic, technical, clinical, and reporting variables from each included publication. Bibliographic fields consisted of publication year, authorship, corresponding-author country, journal, publisher, and keywords. Technical fields included dataset source and size, imaging modality, preprocessing, segmentation method, architecture, training or validation strategy, and reported metrics. Clinical fields recorded the target condition and grouped studies into skin cancer, inflammatory or chronic disease, and infectious disease applications.

The gathered evidence was summarized with a descriptive and thematic approach. Publication numbers, geographic regions, publishing houses, frequently used keywords, model types, data sets, and mentioned limitations were all presented with no statistical analysis. We did not try to conduct a meta-analysis because the ways of disease definitions, sources of images, types of classes, validation sets, and measures of outcome varied so much that there was no possible way of reliably grouping the studies. Thus, accuracy results mentioned in the studies should be considered purely qualitative.

Reporting quality was evaluated based on five criteria: transparency of the dataset (QA1), clarity of preprocessing (QA2), description of the architecture and training strategy (QA3), breadth of evaluation measures (QA4), and acknowledgment of limitations or bias (QA5). When the criterion was completely reported, it got 1 point; if partially reported, 0.5; and 0 if missing. Total scores ranged from 0 to 5 and were classified as high (4.0, 5.0), moderate (3.0, 3.9), or low (<3.0). This is a completeness-of-reporting checklist; it is neither a validated diagnostic risk-of-bias tool nor does it establish clinical validity. **Table 3** presents reporting-quality criteria applied to the included studies.

**Table 3. Reporting-quality criteria applied to the included studies.**

| Code | Criterion                  | Assessment focus                                                                                           |
|------|----------------------------|------------------------------------------------------------------------------------------------------------|
| QA1  | Dataset transparency       | Dataset source, composition, sample size, and class structure are reported.                                |
| QA2  | Preprocessing clarity      | Image resizing, normalization, enhancement, augmentation, and artefact handling are described.             |
| QA3  | Architectural transparency | Architecture, transfer-learning strategy, training procedure, and key implementation choices are reported. |
| QA4  | Metric diversity           | Multiple task-appropriate metrics are provided rather than accuracy alone.                                 |
| QA5  | Limitation analysis        | Limitations, bias, class imbalance, generalizability, or computational constraints are discussed.          |

Applying the five-level checklist identified 36 papers (90.0%) as exemplary in reporting quality and only 4 (10.0%) as average. No articles scored lower than 3.0. Papers scoring in the moderate category mostly reflected the lack of detailed information on how preprocessing, implementation, metrics, and limitations were addressed.

## 2.5 Methodological Scope and Limitations

A targeted systematic design was employed in this review instead of a statistical meta-analysis. Variability in disease labels, image sources, class prevalence rates, methods of train-validation-test split, and outcomes reporting hindered the possibility of any meaningful data pooling. Hence, performance metrics are described in the context of their original studies and should not be taken as an official order of architectural rankings.

In this regard, we only searched four sources, limited only to full-text English-language papers and focused primarily on the years 2017, 2026; moreover, the study was not prospectively registered. The reason is the dynamic character of Google Scholar results, and the exclusion stage is basically the stage of zero- exclusion after deduplication which is a sign of narrow query design. Such features might omit relevant studies and reduce the reproducibility of search results. The study was also criticized for using one of the tools which is an assess of reporting completeness instead of measuring risk of bias.

## 3. EVIDENCE SYNTHESIS

### 3.1 Image Preprocessing

Preparation of the input data is important since a model might have different features available to it depending on things like hair, ruler marks, shadows, changes in colours, compression, or uneven lighting. The pre-processing steps of the reviewed pipelines included combinations of resizing, intensity normalization, denoising, contrast enhancement, colour correction, and artefact removal prior to segmentation or classification [1, 11, 30].

Typically, a series of methods such as morphological filtering and inpainting are combined in the course of hair-removal procedures. On the other hand, median filters and contrast-limited adaptive histogram equalization have been implemented for noise removal or for an increase in local contrast [11, 30]. Not all preprocessing methods bring improvement: over-aggressive enhancement may change diagnostic colour or border details, hence transformations should be defined, tested, and repeated accurately to be reliable.

### 3.2 Lesion Segmentation

By isolating the lesion from adjacent skin, segmentation trims away irrelevant background before classification. Thresholding, clustering, and GrabCut still serve as useful baselines. Their output can deteriorate, however, when contrast is weak, when hair crosses the lesion, or when the border is diffuse [11].

The use of fully convolutional, encoder, decoder, multi-focus, region-based, and ensemble networks has resulted in a more accurate mapping of individual pixels on benchmark datasets [31, 32, 33, 34, 35, 36]. Hybrid CNN, transformer models take this idea one step further by merging local texture modeling with the broader spatial context [14, 17]. Although Dice and Jaccard scores are quite helpful, they do not give the actual clinical meaning which heavily relies on annotation quality and whether errors happen at diagnostically important borders.

### 3.3 Disease Classification

Classification systems these days are trained directly from pixels using either pretrained or task-specific CNN models. VGG, ResNet, DenseNet, Inception and the likes are still popular optimized CNN variants on one hand whereas attention and cost-sensitive objectives try to draw the learning attention towards relevant areas or minority classes [4, 12, 37, 38, 39].

EfficientNet, vision transformers, and hybrid models expand the design space by varying parameter count, receptive field, and data needs [15, 17]. However, since studies differ in labels, image sources, prevalence of classes, and test-set construction, a higher accuracy is insufficient by itself to show that a model is superior or more clinically practical.

### 3.4 Automated Detection Across Dermatological Conditions

Although melanoma and other skin cancers form the main focus of research articles, image-based systems are currently being explored for use in eczema, psoriasis, wider differential diagnosis [1, 8, 40]. It has also been found through human-AI collaboration research that the provision of decision support enhances clinician performance if the model is trustworthy, but can at the same time be very misleading if assistance is erroneous and no matter their level of experience, users are exposed to the possibility of being misled [10].

Transfer learning was also applied for newly appeared infectious diseases such as monkeypox [41]. These results of viability show technical flexibilities, but at the same time, the use of tiny web-downloaded datasets, unreliable labels, and restricted diversity of participants restrict the interpretation of whether the technique is ready for clinical use.

### 3.5 Benchmark Datasets and Performance Evaluation

ISIC, HAM10000, PH2, PAD-UFES-20, and patient-derived melanoma sets allow for reproducible experiments in dermoscopy and clinical photography [18, 20, 22, 23, 24]. Commonly, results are expressed in accuracy, precision, recall, F1-score, AUC (area under the receiver operating characteristic curve), Dice coefficient, or Jaccard index [13].

Rare conditions and very different class sizes make total accuracy especially unreliable. To tackle imbalance Data augmentation, reweighting, resampling, cost-sensitive losses, and attention mechanisms are applied [13, 16, 39]. A much-needed informative evaluation is to present class-wise sensitivity and specificity, confidence intervals, calibration, decision thresholds, and performance by demographic subgroup.

### 3.6 Explainable Artificial Intelligence in Dermatology

For instance, post-hoc explanation techniques Grad-CAM, saliency maps, and attention visualization are methods of pointing out to the pixels or region that was the basis of your prediction [6, 37, 39, 42]. These kinds of explanations might be useful for dermatologists to see if the network pays heed to the lesion or a

surrounding artefact, they can also be used for discussing mistakes and for the clinicians, model conversations.

A heat map does not show cause and effect so one should not see it as a justification of the validity of clinically relevant reasoning done by models. One should check explanations for how truthful they are, what happens when minor image changes happen, that is how sensitive they are model parameters, and whether they are helpful to the targeted clinicians, among other things [42]. The fairness evidence also demonstrates that a good-looking picture will not make up for a badly representative dataset or a performance difference between subgroups [9, 25].

Explanation outputs should be examined, in the context of uncertainty, subgroup performance and failure mode for the clinical deployment. A model that spots a plausible area visually might be poorly calibrated or based on a biased dataset; so, explanation is just one aspect of validation and never a replacement for it [9, 25, 42].

### 3.7 Comparative Analysis of Dermatological AI Approaches

The comparative tables group the studies according not to their performance but to clinical tasks. The synthesized information includes datasets, preprocessing or validation strategies, network architecture, the reported outcome, as well as the main drawback for each study. This representation reveals the diversity of the evidence context and it prevents considering tasks and test sets that are not alike as being directly comparable. **Table 4** presents selected approaches to skin-cancer and melanoma detection or segmentation.

**Table 4. Selected approaches to skin-cancer and melanoma detection or segmentation.**

| Study                      | Dataset        | Preprocessing / segmentation          | Architecture                   | Performance            | Reported limitation                |
|----------------------------|----------------|---------------------------------------|--------------------------------|------------------------|------------------------------------|
| Ahammed et al. [11]        | ISIC; HAM10000 | Black-hat hair removal; GrabCut       | SVM + GLCM                     | 97.0% accuracy         | Manual feature dependence          |
| Al-Masni et al. [31]       | ISIC 2017; PH2 | Full-resolution segmentation          | FrCN                           | 89.3% accuracy         | Boundary-noise sensitivity         |
| Bi et al. [34]             | ISIC 2016      | Multistage segmentation               | FCN                            | 91.18% Dice            | High training cost                 |
| Nida et al. [36]           | ISIC 2016      | Fuzzy C-means                         | Region-based CNN               | 89.0% accuracy         | Weak localization of small lesions |
| Alahmadi and Alghamdi [14] | ISIC 2018      | CNN–transformer feature coupling      | Hybrid CNN–transformer         | Reported improvement   | Large data requirement             |
| Goyal et al. [35]          | ISIC; PH2      | Ensemble segmentation                 | Mask R-CNN + DeepLabV3+        | 89.93% sensitivity     | Annotation intensive               |
| Ravi [39]                  | HAM10000; ISIC | Attention and cost-sensitive learning | Attention fusion network       | 99.0% accuracy         | Complex architecture               |
| Alenezi et al. [43]        | ISIC; PH2      | RANSAC hair removal                   | Multi-task network             | 98.6% accuracy         | Overfitting risk                   |
| Malik et al. [38]          | ISIC Archive   | Lightweight preprocessing             | Optimized CNN                  | 87.64% accuracy        | Below ensemble performance         |
| Groh et al. [9]            | Fitzpatrick17k | Skin-tone labels                      | Deep-learning decision support | Bias analysis          | Skin-tone disparities              |
| Liu et al. [37]            | HAM10000       | Inverted residual blocks              | Attention CNN                  | Multi-class evaluation | Limited explainability             |

Researchers mostly use the same skin-cancer data sets, like ISIC, HAM10000, and PH2, for their skin cancer skin image analysis. These sets are a great way to keep results comparable, but they can be limiting if one just optimizes for the same datasets multiple times as that would result in an insufficient basis of evidence for the method. Prior to general use, high internal performance of a model must be verified by external testing, skin-tone analysis, calibration, workflow evaluation and interpretable error review [9, 16, 19, 21, 25].

**Table 5. Image-based deep-learning evidence for emerging viral skin disease.**

| Study           | Dataset                       | Preprocessing                              | Architecture               | Performance            | Limitation                            |
|-----------------|-------------------------------|--------------------------------------------|----------------------------|------------------------|---------------------------------------|
| Ali et al. [41] | Monkeypox Skin Lesion Dataset | Data augmentation; resizing; normalization | ResNet50 transfer learning | 82.96% ± 4.57 accuracy | Small, demographically narrow dataset |

The monkeypox evidence in **Table 5** is, essentially, just a feasibility study. Transfer learning can generate experimental discrimination based on limited data, but in order to be confident about generalization you need larger, clinically curated cohorts with verified labels and independent testing [41]. **Table 6** presents machine-learning and deep-learning models for inflammatory skin disease.

**Table 6. Machine-learning and deep-learning models for inflammatory skin disease.**

| Study                  | Dataset                   | Preprocessing / validation                     | Model / classifier | Performance     | Limitation                     |
|------------------------|---------------------------|------------------------------------------------|--------------------|-----------------|--------------------------------|
| Bordoloi et al. [44]   | Dermatology dataset       | Median filtering                               | SVM; random forest | 98.8% accuracy  | Hand-crafted features required |
| Rangaswamy et al. [45] | Roboflow; Kaggle          | Normalization                                  | InceptionV3        | 80.88% accuracy | Minority-class weakness        |
| Sari and Keser [46]    | Kaggle dataset; ISIC 2017 | Relief feature selection; five-fold validation | AlexNet + SVM      | 92.10% accuracy | Three disease categories       |
| Hammad et al. [40]     | Kaggle; DermNet           | Image rescaling                                | DermaCare CNN      | 96.20% accuracy | Annotation dependence          |

In fact, research on eczema, psoriasis, and other similar diseases has shown that these methods give fairly accurate results [40, 44, 45, 46]. But there's still a problem with limited numbers of different types of diseases, underrepresentation of minority categories, uncertain labels, and source-dependent image features that limit applicability [1, 3, 16].

Reported limitations were identified as recurring research gaps aimed at linking technical design choices with consequences related to fairness, reproducibility, deployment, and clinical decision-making. **Table 7** presents recurrent research gaps and their implications for clinical translation.

**Table 7. Recurrent research gaps and their implications for clinical translation.**

| Research gap             | Representative studies                                       | Implication                                                               |
|--------------------------|--------------------------------------------------------------|---------------------------------------------------------------------------|
| Skin-tone bias           | Groh et al. [21–9]; Wen et al. [19]                          | Aggregate metrics may conceal unequal error rates across skin tones.      |
| Class imbalance          | Ravi [39]; Khalaf et al. [16]; Sun et al. [13]               | Rare conditions can have poor sensitivity despite high overall accuracy.  |
| Computational complexity | Kshirsagar et al. [27]; Yousaf et al. [17]; Yuan et al. [28] | Memory, latency, and energy demands can prevent point-of-care deployment. |
| Limited explainability   | Jeong et al. [6]; Liu et al. [37]                            | Plausible heatmaps may not be faithful, stable, or useful to clinicians.  |
| Dataset limitations      | Choy et al. [1]; Debelee [3]; Wen et al. [19]                | Benchmark performance may not transfer to other centres or populations.   |

The most recurring issues are demographic bias, class imbalance, resource-heavy architectures, not being able to see model's reasoning very well and having limited external-validation data only. Use of limited dataset collection is also a contributing factor to these issues since it encourages optimizing for one benchmark only. The emphasis therefore needs to change from single-point accuracy boost to data that are diverse, full transparency of methods, models that are compact and efficient, clear explanations of results, and validation studies from multiple centers (13, 14, 18, 20, 25, 28, 44, 46).

## 4. RESULTS AND DISCUSSION

### 4.1 Bibliometric Profile

The corpus of 40 publications was portrayed through these categories year when the study was published, countries of the corresponding authors, publishers, and repeating keyword phrases. The mentioned frequencies represent the sample; they do not depict an absolute picture of all dermatological AI publications.

#### 4.1.1 Publication Trend

The publication year was used to figure out how the selected literature developed throughout the review period. Figure 2 shows the number of papers per year and the proportion that each year's contribution made to the 40 studies included here.

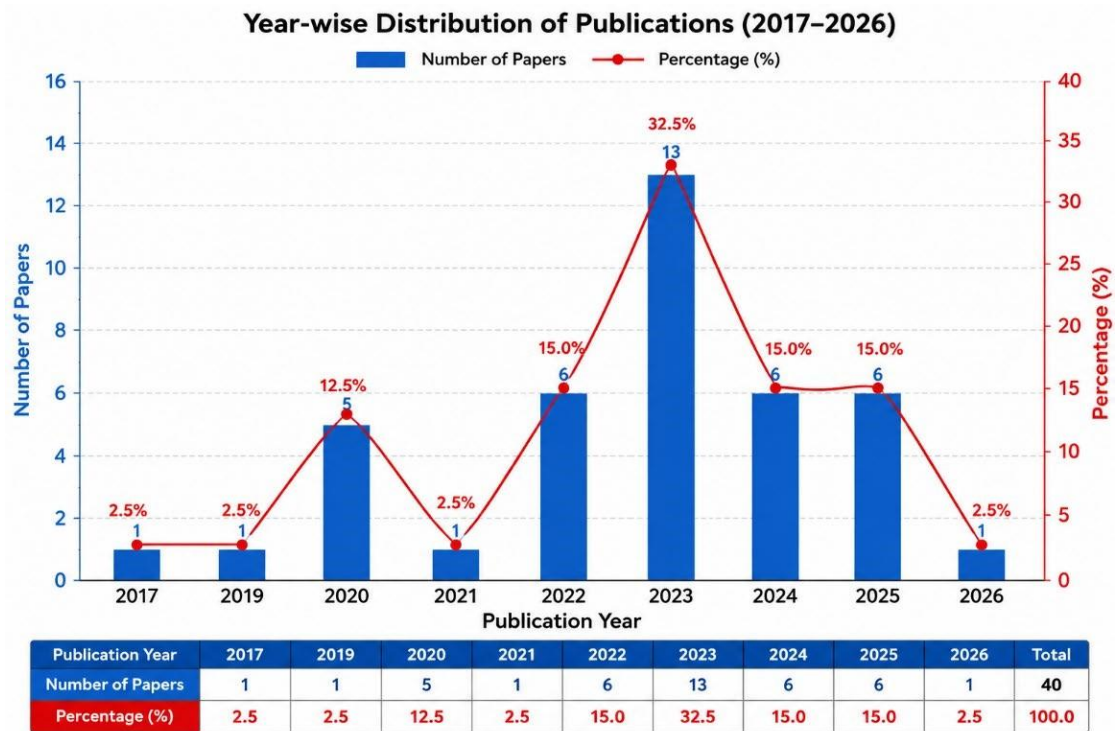


Figure 2. Annual distribution of the included publications (2017–2026), with counts and percentages.

Figure 2 demonstrates a trend wherein the selected corpus got more concentrated post 2021 and it was at maximum level of publications in 2023. One should be very careful when considering the numbers for the past couple of years since the database completeness as well as the timing of the July 2026 search might give the illusion of a downward trend even when it is a publication year that is not complete yet.

#### 4.1.2 Geographical Distribution

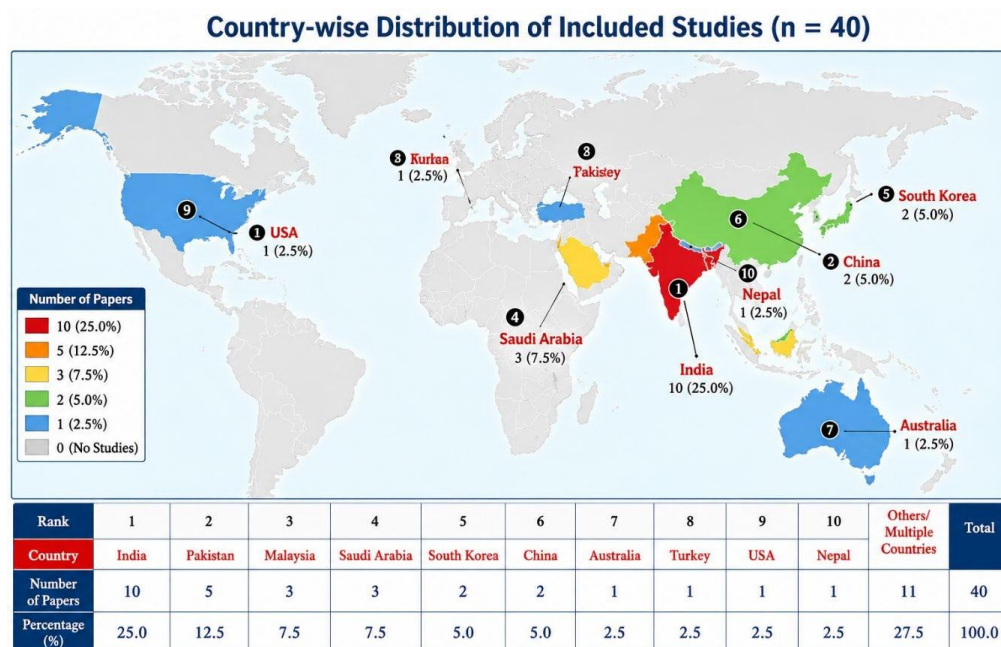


Figure 3. Distribution of included studies by the corresponding author's country.

India and China, along with US, are the leading countries in contribution of corresponding authors from the selected papers as shown in figure 3. The figure shows that there was a major international involvement. However, country counts should not be taken as a measurement of data variety, a

multinational set of authors or publication source does not in itself guarantee representative recruitment of patients.

#### 4.1.3 Publisher Distribution

Identification of publication outlets most represented in the evidence set was carried out by examining publisher counts. This indicator is primarily descriptive and does not reflect the quality of publishers or the whole distribution of the field. It is rather an indicator of a specific sample.

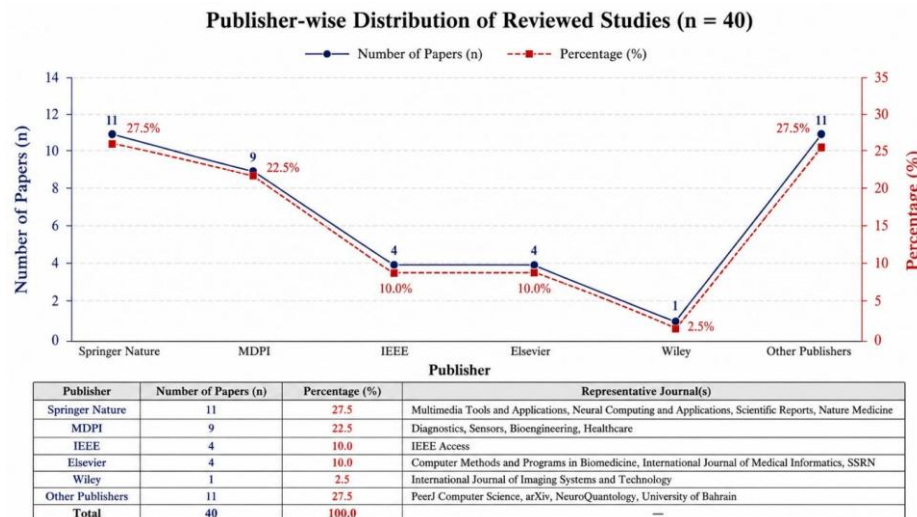


Figure 4. Publisher distribution within the 40-study evidence set.

Springer Nature had the maximum share of included publications, with MDPI and IEEE coming next (Figure 4). Such concentration reflects the interdisciplinary nature of the area, combining clinical medicine, biomedical engineering, and computer vision among others, but it definitely must not be viewed as a publishers' or journals' ranking.

#### 4.1.4 Keyword Profile

The recurring author keywords and indexing terms have been grouped together to reveal a general view of the key topics and trends. Table 8 presents relative prominence of recurring concepts in the included literature.

Table 8. Relative prominence of recurring concepts in the included literature.

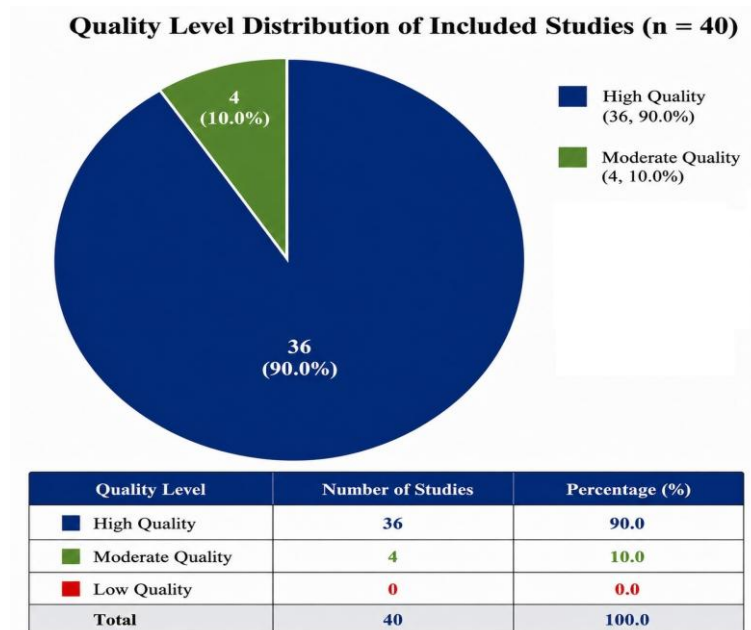
| Keyword                             | Prominence |
|-------------------------------------|------------|
| Deep learning                       | High       |
| Convolutional neural networks       | High       |
| Skin cancer / melanoma              | High       |
| Dermoscopic images                  | High       |
| Image segmentation                  | Moderate   |
| Transfer learning                   | Moderate   |
| Machine learning                    | Moderate   |
| Automated diagnosis                 | Moderate   |
| Feature extraction                  | Moderate   |
| Explainable artificial intelligence | Emerging   |

Deep learning techniques, convolutional neural networks, skin cancer recognition, and dermoscopy are still the main subjects. Segmentation, transfer learning, and automatic diagnosis constitute the second level, and explainable AI is shown as a new topic. It is indicated here that interpretability is gaining more attention; however, the assessment or the explanation of the results is still significantly underdeveloped compared to the reporting of model-performance.

## 4.2 Methodological Quality and Technical Reporting

The technical appraisal looked at how much the different studies described their datasets, preprocessing, architecture, evaluation metrics, and limitations. The emerging categories provide an overview of reporting quality and are to be considered along with the design limitations that have been highlighted in Section 2.5.

#### 4.2.1 Quality-Appraisal Results



**Figure 5.** Distribution of included studies by reporting-quality category.

Out of 36 studies (90.0%), the reporting quality was assessed as high enough, and only 4 (10.0%) were rated as moderate. No one paper fell into the low category (**Figure 5**). Therefore, it is reasonable to assume that most of them contained a significant amount of data for the intended extraction. However, it should not be interpreted as a proof of low-risk bias, external validation success, representative sampling, and medical effectiveness. In fact, a study might report a limited dataset or an imbalanced cohort quite straightforwardly but a high reporting level can still be awarded.

#### 4.3 Interpretation of the Evidence

The publication curve clearly indicates a major spike in image-based dermatological artificial intelligence around 2021 and beyond. The thematic highlight has changed from traditional feature engineering to deep CNNs, attention mechanisms, transformers, and explainability [1, 6, 13]. At the same time, we can see a very strong dependence and limitation of the evidence by benchmark datasets used by the top groups only and retrospective designs.

The technical analysis validates moving beyond melanoma to inflammatory and infectious diseases, where new modelling approaches are introduced by EfficientNet, attention modules, and hybrid CNN, transformer models [15, 17, 37, 40, 41]. Performance differences between studies are a result of more than just architecture since label assignments, disease prevalence, preprocessing steps, data splits, and test patient populations can vary greatly.

In a number of articles, the description of methods remains sufficient for the readers while, still, some crucial deficiencies are recurring like, for instance, imbalance of classes, lack of representation of people of dark skin, the need for high computing resource, doubts about the correctness of the calibration and low external validation [1, 25, 21-9, 16, 19]. Such issues make people question how benchmark results will translate into everyday hospital work. Clinically, there is a need to evaluate the tool according to its actual and intended use case. The clinical-informatics triage tools need to be accurate and safe to refer while differentially diagnosed-diagnosis aids need to produce ranked outputs correctly and clinically-facing systems should be tested for usability & human factors. Evidence from extensive differential diagnosis and human-computer collaboration suggests that workflow design can be as critical as standalone model accuracy [8, 10]. **Table 9** presents priority directions for research and clinical translation.

**Table 9. Priority directions for research and clinical translation.**

| s                           | Primary objective                                     | Expected contribution                            |
|-----------------------------|-------------------------------------------------------|--------------------------------------------------|
| Clinically validated XAI    | Test fidelity, stability, and clinician usefulness    | Safer interpretation and auditable decisions     |
| Vision transformers         | Evaluate global context without benchmark overfitting | Evidence on when transformers add clinical value |
| Lightweight edge AI         | Reduce latency, memory, and power demand              | Feasible mobile and low-resource deployment      |
| Federated learning          | Support multi-centre learning with privacy controls   | Broader data without routine centralization      |
| Skin-tone-diverse datasets  | Measure and reduce demographic performance gaps       | Fairer and more transportable systems            |
| Multimodal decision support | Combine images with history and clinical metadata     | Predictions closer to real diagnostic reasoning  |

The next phase of dermatological AI should be measured through clinicians' meaningful feedback and not only accuracy [9, 10, 16, 17, 28].

Recent reviews highlight two additional directions relevant to dermatology AI. Hybrid feature-fusion architectures that couple complementary backbones have shown that combined representations can improve discrimination under data scarcity [47]. Similarly, advanced neural computation is increasingly reviewed as a cross-cutting capability for clinical image diagnostics across modalities [48].

## 5. CONCLUSION

The review draws on and incorporates evidence relating to image preprocessing, lesion segmentation, disease classification, explainability, and clinical translation of dermatological AI. CNNs still make up the primary model family. Besides CNNs, EfficientNet, attention, vision transformers, and hybrid approaches enrich the list of methods. Public datasets such as ISIC, HAM10000, PH2, PAD-UFES-20, and melanoma databases based on patient data have contributed immensely toward facilitating the development and comparison.

In conclusion, it is a conditional statement: a very successful deep-learning system on one image task does not mean that such systems could easily be brought into routine usage in a doctor's clinic or hospital based on benchmark results and accuracy alone. Problems like demographic imbalance, class imbalance, artefacts, computational demands, limited external testing, weak calibration, and unvalidated explanations continue to be significant barriers.

Future works are expected to include diverse multicenter imaging along with corresponding clinical context, assess performance in different skin tones and rare classes, disclose uncertainty and calibration, and determine if explanations lead to better decision-making or merely look appealing. An efficient network architecture and a human-AI interaction designed in a way that considers both human and machine capabilities may help increase access, but without a prospective verification, good governance, and ongoing performance checks these may not be enough. These measures are required in order for dermatologic decision aid to be safe, fair, and responsible.

## AUTHOR CONTRIBUTIONS

Dr Satish Kumar Sahu was responsible for giving academic direction to the research. Abhishek Tiwari has done the job of literature search, data extraction, evidence synthesis, and manuscript preparation. Both the authors have critically reviewed and approved the manuscript.

## ETHICS STATEMENT

This is a systematic review that collects results from prior articles but does not conduct new research with human or animal subjects, for this reason, institutional ethics approval together with informed consent were not necessary.

## DATA AVAILABILITY

The entire review relies only on information from papers cited in this reference section. The organized body of evidence together with key findings are presented in this paper and illustrated in the corresponding charts and tables.

## FUNDING

The authors have not been supported by any specific grant from a public, commercial or not-for-profit funding agency for this work.

## CONFLICT OF INTEREST

The authors have made a declaration that they do not have any competing interests.

## REFERENCES

- [1] Choy, S. P., Kim, B. J., Paolino, A., Tan, W. R., Lim, S. M. L., Seo, J., Tan, S. P., Francis, L., Tsakok, T., Simpson, M., Barker, J. N. W. N., Lynch, M. D., Corbett, M. S., Smith, C. H., & Mahil, S. K. (2023). Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. *npj Digital Medicine*, 6, 180. <https://doi.org/10.1038/s41746-023-00914-8>
- [2] Trakatelli, M., Richard, M.-A., Rouillard, A., Paul, C., Röcken, M., & Stratigos, A. (2023). The burden of skin disease in Europe. *Journal of the European Academy of Dermatology and Venereology*, 37(Suppl. 7), 3–5. <https://doi.org/10.1111/jdv.19390>
- [3] Debelee, T. G. (2023). Skin lesion classification and detection using machine learning techniques: A systematic review. *Diagnostics*, 13(19), 3147. <https://doi.org/10.3390/diagnostics13193147>
- [4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [5] Mazhar, T., Haq, I., Ditta, A., Mohsan, S. A. H., Rehman, F., Zafar, I., Gansau, J. A., & Goh, L. P. W. (2023). The role of machine learning and deep learning approaches for the detection of skin cancer. *Healthcare*, 11(3), 415. <https://doi.org/10.3390/healthcare11030415>
- [6] Jeong, H. K., Park, C., Henao, R., & Kheterpal, M. (2023). Deep learning in dermatology: A systematic review of current approaches, outcomes, and limitations. *JID Innovations*, 3(1), 100150. <https://doi.org/10.1016/j.xjidi.2022.100150>
- [7] Kawahara, J., & Hamarneh, G. (2019). Visual diagnosis of dermatological disorders: Human and machine performance [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.1906.01256>
- [8] Liu, Y., Jain, A., Eng, C., Way, D. H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., Gupta, V., Singh, N., Natarajan, V., Hofmann-Wellenhof, R., Corrado, G. S., Peng, L. H., Webster, D. R., Ai, D., Huang, S. J., ... Coz, D. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, 26(6), 900–908. <https://doi.org/10.1038/s41591-020-0842-3>
- [9] Groh, M., Badri, O., Daneshjou, R., Koochek, A., Harris, C., Soenksen, L. R., Doraiswamy, P. M., & Picard, R. (2024). Deep learning-aided decision support for diagnosis of skin disease across skin tones. *Nature Medicine*, 30(2), 573–583. <https://doi.org/10.1038/s41591-023-02728-3>
- [10] Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8), 1229–1234. <https://doi.org/10.1038/s41591-020-0942-0>
- [11] Ahammed, M., Mamun, M. A., & Uddin, M. S. (2022). A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*, 2, 100122. <https://doi.org/10.1016/j.health.2022.100122>

- [12] Bandyopadhyay, S. K., Bose, P., Bhaumik, A., & Poddar, S. (2022). Machine learning and deep learning integration for skin diseases prediction. *International Journal of Engineering Trends and Technology*, 70(2), 11–18. <https://doi.org/10.14445/22315381/IJETT-V70I2P202>
- [13] Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., & Yang, X. (2023). Machine learning methods in skin disease recognition: A systematic review. *Processes*, 11(4), 1003. <https://doi.org/10.3390/pr11041003>
- [14] Alahmadi, M. D., & Alghamdi, W. (2022). Semi-supervised skin lesion segmentation with coupling CNN and transformer features. *IEEE Access*, 10, 122560–122569. <https://doi.org/10.1109/ACCESS.2022.3224005>
- [15] Fatima, S., Akram, M. U., Mohammad, S., & Ahmed, S. B. (2025). Deep learning in dermatopathology: Applications for skin disease diagnosis and classification. *Discover Applied Sciences*, 7(9), 1006. <https://doi.org/10.1007/s42452-025-07138-3>
- [16] Khalaf, A. D., Hamdan, H., Abdul Halin, A., & Manshor, N. (2025). Segmentation and classification of skin cancer diseases based on deep learning: Challenges and future directions. *IEEE Access*, 13, 90163–90184. <https://doi.org/10.1109/ACCESS.2025.3569170>
- [17] Yousaf, N., Amin, J., Butt, W. H., Zafar, A., & Kim, S. (2026). Advanced hybrid transformer–CNN framework for improved skin lesion classification and segmentation. *Scientific Reports*, 16, 13592. <https://doi.org/10.1038/s41598-026-43376-0>
- [18] Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kallou, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging, hosted by the International Skin Imaging Collaboration. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168–172). IEEE. <https://doi.org/10.1109/ISBI.2018.8363547>
- [19] 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., & Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: A systematic review. *The Lancet Digital Health*, 4(1), e64–e74. [https://doi.org/10.1016/S2589-7500\(21\)00252-1](https://doi.org/10.1016/S2589-7500(21)00252-1)
- [20] Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Liopyris, K., Malvey, J., Musthaq, S., Nanda, J., ... Soyer, H. P. (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, 8, 34. <https://doi.org/10.1038/s41597-021-00815-z>
- [21] Groh, M., Harris, C., Soenksen, L. R., Lau, F., Han, R., Kim, A., Koochek, A., & Badri, O. (2021). Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1820–1828). IEEE. <https://doi.org/10.1109/CVPRW53098.2021.00201>
- [22] Mendonça, T., Ferreira, P. M., Marques, J. S., Marçal, A. R. S., & Rozeira, J. (2013). PH2—A dermoscopic image database for research and benchmarking. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5437–5440). IEEE. <https://doi.org/10.1109/EMBC.2013.6610779>
- [23] Pacheco, A. G. C., Lima, G. R., Salomão, A. S., Krohling, B., Biral, I. P., de Angelo, G. G., Alves, F. C. R., Jr., Esgario, J. G. M., Simora, A. C., Castro, P. B. C., Rodrigues, F. B., Frasson, P. H. L., Krohling, R. A., Knidel, H., Santos, M. C. S., do Espírito Santo, R. B., Macedo, T. L. S. G., Canuto, T. R. P., & de Barros, L. F. S. (2020). PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, 32, 106221. <https://doi.org/10.1016/j.dib.2020.106221>
- [24] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. <https://doi.org/10.1038/sdata.2018.161>
- [25] Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Mukherjee, P., Phung, M., Yekrang, K., Fong, B., Sahasrabudhe, R., Allerup, J. A. C., Okata-Karigane, U., Zou, J., & Chiou, A. S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*, 8(32), eabq6147. <https://doi.org/10.1126/sciadv.abq6147>
- [26] Groh, M., Harris, C., Daneshjou, R., Badri, O., & Koochek, A. (2022). Towards transparency in dermatology image datasets with skin tone annotations by experts, crowds, and an algorithm [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2207.02942>
- [27] Kshirsagar, P. R., Manoharan, H., Shitharth, S., Alshareef, A. M., Albishry, N., & Balachandran, P. K. (2022). Deep learning approaches for prognosis of automated skin disease. *Life*, 12(3), 426. <https://doi.org/10.3390/life12030426>
- [28] Yuan, C., Zhao, D., & Agaian, S. S. (2024). UCM-Net: A lightweight and efficient solution for skin lesion segmentation using MLP and CNN. *Biomedical Signal Processing and Control*, 96, 106573. <https://doi.org/10.1016/j.bspc.2024.106573>

- [29] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- [30] Haner Kırğıl, E. N., & Erdaş, Ç. B. (2024). Enhancing skin disease diagnosis through deep learning: A comprehensive study on dermoscopic image preprocessing and classification. *International Journal of Imaging Systems and Technology*, 34(4), e23148. <https://doi.org/10.1002/ima.23148>
- [31] Al-Masni, M. A., Kim, D.-H., & Kim, T.-S. (2020). Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Computer Methods and Programs in Biomedicine*, 190, 105351. <https://doi.org/10.1016/j.cmpb.2020.105351>
- [32] Anjum, M. A., Amin, J., Sharif, M., Khan, H. U., Malik, M. S. A., & Kadry, S. (2020). Deep semantic segmentation and multi-class skin lesion classification based on convolutional neural network. *IEEE Access*, 8, 129668–129678. <https://doi.org/10.1109/ACCESS.2020.3009276>
- [33] Basak, H., Kundu, R., & Sarkar, R. (2022). MFSNet: A multi-focus segmentation network for skin lesion segmentation. *Pattern Recognition*, 128, 108673. <https://doi.org/10.1016/j.patcog.2022.108673>
- [34] Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., & Feng, D. (2017). Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, 64(9), 2065–2074. <https://doi.org/10.1109/TBME.2017.2712771>
- [35] Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2020). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*, 8, 4171–4181. <https://doi.org/10.1109/ACCESS.2019.2960504>
- [36] Nida, N., Irtaza, A., Javed, A., Yousaf, M. H., & Mahmood, M. T. (2019). Melanoma lesion detection and segmentation using deep region-based convolutional neural network and fuzzy C-means clustering. *International Journal of Medical Informatics*, 124, 37–48. <https://doi.org/10.1016/j.ijmedinf.2019.01.005>
- [37] Liu, H., Dou, Y., Wang, K., Zou, Y., Sen, G., Liu, X., & Li, H. (2025). A skin disease classification model based on multi-scale combined efficient channel attention module. *Scientific Reports*, 15, 6116. <https://doi.org/10.1038/s41598-025-90418-0>
- [38] Malik, S. G., Jamil, S. S., Aziz, A., Ullah, S., Ullah, I., & Abohashrh, M. (2024). High-precision skin disease diagnosis through deep learning on dermoscopic images. *Bioengineering*, 11(9), 867. <https://doi.org/10.3390/bioengineering11090867>
- [39] Ravi, V. (2022). Attention cost-sensitive deep learning-based approach for skin cancer detection and classification. *Cancers*, 14(23), 5872. <https://doi.org/10.3390/cancers14235872>
- [40] Hammad, M., Pławiak, P., ElAffendi, M., El-Latif, A. A. A., & Latif, A. A. A. (2023). Enhanced deep learning approach for accurate eczema and psoriasis skin detection. *Sensors*, 23(16), 7295. <https://doi.org/10.3390/s23167295>
- [41] Ali, S. N., Ahmed, M. T., Paul, J., Jahan, T., Sani, S. M. S., Noor, N., & Hasan, T. (2022). Monkeypox skin lesion detection using deep learning models: A feasibility study [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2207.03342>
- [42] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- [43] Alenezi, F., Armghan, A., & Polat, K. (2023). A novel multi-task learning network based on melanoma segmentation and classification with skin lesion images. *Diagnostics*, 13(2), 262. <https://doi.org/10.3390/diagnostics13020262>
- [44] Bordoloi, D., Singh, V., Kaliyaperumal, K., Ritonga, M., Jawarneh, M., Kassanuk, T., & Quiñonez-Choquecota, J. (2023). Classification and detection of skin disease based on machine learning and image processing evolutionary models. *Computer Assisted Methods in Engineering and Science*, 30(2), 247–256. <https://doi.org/10.24423/CAMES.479>
- [45] Rangaswamy, S., Tantry, S. S., & Lal, T. S. (2025). Skin disease classification using deep learning. *National Academy Science Letters*, 48(5), 585–588. <https://doi.org/10.1007/s40009-024-01523-z>
- [46] Sari, M. O., & Keser, K. (2025). Classification of skin diseases with deep learning-based approaches. *Scientific Reports*, 15, 27506. <https://doi.org/10.1038/s41598-025-13275-x>
- [47] Singh, S., & Pandey, J. K. (2026). Hybrid deep feature fusion for facial emotion recognition using VGG19 and ResNet152V2. *International Journal of Innovative Research in Computer Science and Technology*, 14(3), 17–26. <https://doi.org/10.55524/ijircst.2026.14.2.10>

- [48] Pandey, J. K., Verma, S. K., Kumar, J., Perwej, Y., Jha, S. K., Panchal, B. Y., Ferdouse, R., Sindhu, V., Banerjee, S., Tiwari, M., Badal, R., Mandal, P., & Baghel, J. S. (2026). Transformative role of advanced neural computation in clinical image diagnostics: A review of key concepts and applications. *Seminars in Ultrasound, CT and MRI*. Advance online publication. <https://doi.org/10.1053/j.sult.2026.06.010>