Anadolu Psikiyatri Dergisi || ISSN: 1302-6631 || Volume 27; Number 1
https://doi.org/10.5281/zenodo.18383394

The Role of Artificial Intelligence in Modern Dermatological Diagnostics

Authors: Kornella Zieleń[1] ORCID: 0009- 0007-3440-909X, Izabela Majchrzak[1] ORCID: 0009-0009-0682-7184, Anna Lemańska[3] ORCID: 0009-0008-3207-1248,Martyna Lenartowicz[1] ORCID: 0009-0008-3452-6090, Dagmara Kanar[4] ORCID: 0009-0005-5577-5334, Sandra Moska[2] ORCID: 0009-0004-5475-3216, Joanna Cygan[5] ORCID: 0009-0000-5512-8261

Authors' Affiliation:

[1]University Clinical Hospital in Poznań, Poland

[2]Regional Hospital in Poznań, Poland

[3]Central Clinical Hospital, Uniwersity Clinical Center of the Medical University of Warsaw, Poland

[4]Collegium Medicum, Cardinal Stefan Wyszyński University in Warsaw, Faculty of Medicine, Poland

[5]Independent Public Health Care Facility of the Ministry of Interior and Administration in Katowice, Poland

Corresponding Author: Kornella Zieleń

[1]University Clinical Hospital in Poznań, Poland

Abstract

Artificial intelligence (AI) has become an increasingly important tool in dermatology, largely due to the visual nature of skin disease assessment and the rapid development of machine learning techniques. In particular, deep learning models based on convolutional neural networks have demonstrated high performance in the analysis of clinical and dermoscopic images, enabling automated classification, segmentation and detection of skin lesions. Numerous studies report diagnostic accuracy comparable to that of experienced dermatologists in controlled experimental settings. Recent research extends beyond retrospective validation and emphasizes clinical implementation, human–AI collaboration and integration into diagnostic workflows, including total body photography and longitudinal surveillance of high-risk patients. Despite these advances, significant challenges persist, such as limited model generalizability, underrepresentation of darker skin phototypes, variability in image acquisition conditions and the limited interpretability of deep learning systems. This review summarizes current applications of AI in dermatology, highlights key limitations and outlines future directions for the safe, transparent and equitable integration of AI into routine clinical practice.

1. Introduction

Artificial intelligence (AI) is a field of computer science focused on the development of systems capable of performing tasks that traditionally require human intelligence, such as pattern recognition, classification, prediction and decision support. In medical applications, AI primarily relies on machine learning and deep learning algorithms, which enable systems to learn from large datasets and to identify complex, non-linear relationships within data [1].

Dermatology represents a particularly suitable domain for the application of artificial intelligence due to its strong reliance on visual assessment. The diagnostic process in dermatology is largely based on the evaluation of skin lesions, dermoscopic structures and morphological patterns, which can be effectively analysed using image-based deep learning models, especially convolutional neural networks (CNNs) [1]. Advances in these methods have enabled the development of computer-aided diagnostic systems capable of classifying and segmenting skin lesions with performance approaching that of experienced dermatologists under controlled conditions [7-12].

The increasing interest in AI in dermatology is driven by clinical challenges such as interobserver variability, the growing burden of skin cancer surveillance and limited access to specialist care in some healthcare settings. [8,10-13] Large, publicly available image databases, including the International Skin Imaging Collaboration (ISIC) archive, HAM10000 and Derm7pt datasets, have played a key role in facilitating the training and validation of dermatological AI models [2–6].

Despite these advances, significant limitations remain, including restricted generalizability of models trained on homogeneous datasets, underrepresentation of diverse skin phototypes and limited transparency of deep learning decision processes [27-29]. Consequently, artificial intelligence should be regarded as a supportive tool designed to augment clinical judgement rather than replace it, with its safe implementation requiring robust validation and careful integration into clinical workflows. [13,18,27-29].

2. FOUNDATIONS OF ARTIFICIAL INTELLIGENCE IN DERMATOLOGY

Artificial intelligence (AI) has become an asset in dermatology largely because the specialty heavily depends on visual evaluation of skin lesions, rashes and dermatoscopic features. Among the AI techniques, deep learning, especially convolutional neural networks (CNNs) stands out as the most commonly utilized method for examining dermatological images. CNNs facilitate the automated analysis of collections of clinical and dermoscopic images, aiding in processes like image classification, segmentation and lesion identification, with accuracy approaching that of expert dermatologists. [1]

2.1 Convolutional Neural Networks (CNNs)

CNNs are a class of deep neural networks specifically designed to process image data. Their architecture incorporates convolutional layers that automatically learn hierarchical image features. Early layers detect basic structures such as edges, textures, and color gradients, while deeper layers capture more complex visual patterns relevant to dermatological diagnosis, such as pigment networks, vascular structures, or lesion asymmetry. [1]

2.1.1. Input Layer

The input layer receives raw pixel data, typically represented as a 2D or 3D tensor. Unlike traditional neural networks, which flatten images into vectors, CNNs preserve spatial structure, allowing the model to detect patterns based on location and proximity of pixels. [1]

2.1.2 Convolutional Layers

These layers apply small filters (kernels) that slide across the image to detect local features such as edges, textures, or changes in color. As layers deepen, CNNs learn increasingly complex patterns relevant to dermatology, such as lesion borders or pigment structures. [1]

2.1.3 Activation Functions

Following each convolution, activation functions such as ReLU (Rectified Linear Unit) introduce nonlinearity into the network. Without nonlinear activation, the network would behave like a linear filter and fail to learn complex visual patterns.[1]

2.1.4 Pooling Layers

Pooling reduces the spatial size of feature maps while keeping the most important information. The most frequently used method is max pooling. This step decreases computation and enhances the model's resilience to minor variations in the location or appearance. [1]

2.1.5. Fully Connected Layers

Toward the end of the CNN, feature maps are flattened and passed to fully connected (dense) layers. These layers combine the extracted features to perform decision-making tasks such as:

- classifying lesions (e.g., benign vs. malignant),
- predicting disease categories,
- generating risk scores.

2.1.6. Training Process

During training, the weights of the CNN are adjusted using backpropagation along with gradient descent. Techniques like data augmentation and dropout help prevent overfitting and improve generalization. [1]

2.2. Commonly Used AI Techniques in Dermatological Imaging

2.2.1. Image Classification

Image classification is the most widely studied task in dermatological AI. The goal is to sort an image into a specific diagnostic category, which may include differentiating between malignant and benign lesions or recognizing conditions like eczema, psoriasis, and vitiligo.

CNN-based classifiers are typically trained on large datasets of labeled images, using loss functions such as categorical cross-entropy. Their effectiveness is measured with indicators like accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). Importantly, models for identifying melanoma have shown levels of accuracy similar to that of board-certified dermatologists in controlled experimental environments. [1]

2.2.2. Image Segmentation

Segmentation involves identifying the edges of skin lesions within image. This process is crucial for measuring lesion asymmetry, color distribution, and irregular borders—key diagnostic features in dermoscopy. Segmentation improves downstream tasks by:

- isolating relevant lesion areas,
- reducing noise from surrounding skin,
- enabling pixel-level analysis for malignancy prediction.

Applications include melanoma border extraction, psoriasis plaque quantification, and automated scoring of inflammatory skin diseases. [1]

2.2.3. Object Detection

Object detection algorithms identify and localize one or more lesions within a broader field of view. This is particularly relevant for total-body photography, where patients may present with dozens or hundreds of nevi requiring surveillance.
Detection systems support doctors by highlighting atypical lesions that need further examination or follow-up. [1]

2.3Transfer Learning in Dermatology

A major challenge in dermatological AI is the limited availability of large, diverse, and well-annotated image datasets. To address this, researchers frequently employ transfer learning, a method in which CNNs pretrained on large general-purpose datasets are fine-tuned on dermatology-specific images.

Advantages of transfer learning include:

- Improved performance with small datasets, which is common in rare dermatoses.
- Faster convergence during training due to initialized feature representations.
- Reduced computational cost, since only the final layers must be retrained.

Transfer learning has become a standard practice in the development of dermatology AI models, enabling high diagnostic accuracy even when available training data are limited or imbalanced. [1]

3. Dermatological Databases Used in the Development of Artificial Intelligence Systems

Recent advances in machine learning have substantially accelerated the development of artificial intelligence (AI)–based algorithms for medical image analysis. In dermatology, AI-driven computer-aided diagnostic (CAD) systems have demonstrated significant potential in the detection and classification of skin lesions, including malignant melanoma and other forms of skin cancer. The principal advantages of AI-based diagnostic tools include their wide accessibility, scalability, and the ability to provide decision support comparable to expert-level assessment, particularly in settings with limited access to experienced dermatologists.

The availability of large, well-annotated, and publicly accessible dermatological image datasets is a fundamental prerequisite for the training, validation, and benchmarking of AI algorithms. Among the most widely used and scientifically validated datasets in this field are the International Skin Imaging Collaboration (ISIC) archive, HAM10000, and Derm7pt.

3.1 International Skin Imaging Collaboration (ISIC)

The International Skin Imaging Collaboration (ISIC) is a global research initiative aimed at reducing skin cancer–related mortality through the advancement of digital skin imaging and artificial intelligence. ISIC integrates multiple dermatological datasets and currently comprises more than 549,000 skin lesion images, making it the largest publicly available repository of its kind to date.

A central component of the initiative is the ISIC Archive, an open-source platform that enables the contribution and dissemination of skin lesion images. The archive primarily focuses on dermoscopic images of individual skin lesions, which are inherently standardized due to the use of specialized acquisition devices and are associated with fewer privacy concerns compared with conventional clinical photographs. Each image is linked to a ground-truth diagnosis, often supported by histopathological confirmation, and may include additional clinical metadata. The ISIC Archive serves as a critical resource for education, research, and algorithm development. It allows researchers to query and download images individually or in batches using faceted search tools. Furthermore, ISIC actively fosters collaboration between the dermatology and computer science communities by organizing scientific meetings, publishing research findings, and hosting internationally recognized AI Grand Challenges, which provide standardized benchmarks for evaluating the performance of diagnostic algorithms.[2]

3.2 HAM10000 Dataset

In 2018, the HAM10000 dataset (“Human Against Machine with 10,000 training images”) was introduced as a large, curated benchmark dataset for skin lesion classification. The dataset consists of 10,015 dermoscopic images collected over a period of approximately 20 years from multiple clinical centers. It includes images representing seven common diagnostic categories of pigmented skin lesions, with diagnoses established through histopathology, follow-up examinations, or expert consensus.[3]

HAM10000 was specifically designed to support academic research in machine learning and is publicly available through the ISIC Archive. Due to its relatively balanced class distribution and standardized image quality, HAM10000 has become one of the most frequently used datasets for training and evaluating convolutional neural networks (CNNs). It is also commonly employed in studies comparing the diagnostic performance of AI systems with that of human experts, thereby providing an important reference point for assessing clinical applicability.[4]

3.3 Derm7pt Dataset

The Derm7pt dataset comprises approximately 2,000 clinical and dermoscopic color images of skin lesions, accompanied by structured clinical information. This dataset was developed to support the training and evaluation of CAD systems based on the seven-point checklist for melanoma, a well-established diagnostic algorithm used in dermoscopic practice.[5] Unlike many image-only datasets, Derm7pt provides multimodal data, including both visual information and annotated clinical criteria related to the seven-point malignancy score. This characteristic has led to widespread use in studies evaluating multitask and multimodal deep learning architectures. Research has demonstrated that models trained on Derm7pt can achieve strong classification performance while maintaining relatively low computational complexity.

The release of the Derm7pt dataset has significantly contributed to the advancement of multimodal learning approaches in dermatology. By enabling the integration of image-based features with structured clinical descriptors, Derm7pt has inspired new methodological strategies for improving the accuracy and interpretability of AI-based skin cancer classification systems, and it continues to serve as a valuable benchmark in this research domain.[6]

4. ARTIFICIAL INTELLIGENCE IN DERMOSCOPY AND TOTAL BODY PHOTOGRAPHY

4.1 Introduction

Artificial intelligence has emerged as a promising tool in the image-based diagnosis of pigmented skin lesions, particularly melanoma, with several landmark studies providing evidence for its potential clinical relevance. In a pivotal study, Esteva et al. (2017) showed that a deep convolutional neural network trained on a large and heterogeneous dataset of 129,450 clinical images encompassing 2,032 skin disease categories achieved diagnostic performance comparable to that of 21 board-certified dermatologists when evaluated on biopsy-proven images in binary classification tasks [7].

Subsequent comparative reader studies extended these findings by directly evaluating the performance of deep learning algorithms against large cohorts of dermatologists under standardized dermoscopic conditions. Using a standardized set of dermoscopic images, Haenssle et al. (2018) reported that the deep learning model demonstrated a higher mean area under the receiver operating characteristic curve than the dermatologist cohort and was associated with higher specificity at matched sensitivity levels. Importantly, dermatologist performance improved when additional clinical information beyond dermoscopic images was incorporated. [8] This observation indicates that, while deep learning models perform strongly on image based assessment alone, human experts benefit from the integration of broader clinical context in melanoma diagnostics.

Rather than focusing solely on performance comparisons, more recent work examined how decision support based on artificial intelligence influences clinical decision making. In this context, Tschandl et al. (2020) investigated the impact of artificial intelligence based decision support on diagnostic accuracy in the assessment of pigmented skin lesions by clinicians with varying levels of experience. The authors demonstrated that appropriately designed artificial intelligence based decision support improved diagnostic accuracy compared with unaided clinician assessment, with the greatest benefit observed among less experienced readers. At the same time, the study highlighted that misleading artificial intelligence outputs could negatively influence clinical decisions, underscoring the risk of overreliance on algorithmic support [9].

4.2 Why this topic matters

Early detection of melanoma and other skin cancers depends on whether we can identify, in time, a lesion that differs from the patient's usual pattern. Dermoscopy improves diagnostic accuracy, but several challenges remain in daily practice: variation between clinicians, a high number of lesions in high-risk patients, and the difficulty of noticing subtle changes over time. For this reason, research on artificial intelligence (AI) and machine learning (ML) tools has developed along two main lines: AI for dermoscopic image analysis and automated support in total body photography (TBP), especially 3D-TBP, where whole-skin mapping and longitudinal comparison are central. [10-24]

4.3 AI in dermoscopy: from image classification to clinical decision support

One of the papers that highlighted the potential of deep learning for skin lesion classification was the study by Esteva et al. in *Nature*, which showed that a neural network can reach very high performance in classifying images of skin lesions. This was an important step because the field moved from a technical curiosity to a realistic direction for medical imaging in dermatology. [10]

Subsequent studies focused specifically on dermoscopy and direct comparisons with physicians. Haenssle et al. compared a convolutional neural network with dermatologists in identifying melanoma on dermoscopic images. These results suggest that the algorithm can perform at a comparable level and, in some settings, may outperform some clinicians. [11] A similar message comes from Brinker et al., who conducted a large head-to-head comparison in which the algorithm achieved strong diagnostic performance on dermoscopic images against a large group of dermatologists. [12]

In clinical practice, the key question is not whether a model "beats" a clinician, but whether patient outcomes improve when clinicians are supported. Tschandl et al., in *Nature Medicine,* described an approach based on human-computer collaboration for skin cancer recognition. Such a setup can improve diagnostic accuracy, but it requires careful consideration of how results are presented and how they influence clinical judgement. [13] This broader issue is also addressed in the meta-analysis by Krakowski et al., which focuses on human-AI interaction and shows that what matters is not only model performance, but also how decision support changes diagnostic behaviour. [18]

At the level of evidence synthesis, Salinas et al. conducted a systematic review and meta-analysis comparing AI systems with clinicians in skin cancer diagnosis. These studies help organise the literature, while also showing that comparisons are difficult because studies differ in datasets, endpoints, patient selection, and testing conditions. [19]

4.4 Prospective studies: when tools are tested beyond curated image sets

The greatest clinical value appears when a tool is assessed in prospective studies under conditions close to routine care. In the study by Menzies et al., clinicians were compared with a mobile phone-based AI system for diagnosis and management of pigmented skin cancer in secondary care. This type of trial is important because it shows how a tool performs in real patient pathways, where the task is not only classification but also decisions about next steps. [14] Marchetti et al. performed a prospective validation of an open-source dermoscopy-based AI system for melanoma diagnosis (PROVE-AI), presenting an approach that emphasises clinical usefulness rather than performance in a purely experimental setting. [15]

A similar direction is seen in the study by Heinlein et al., a prospective multicentre investigation in which AI was used to improve dermoscopic melanoma diagnosis in patient care. The multicentre design matters because it reduces the risk that results are driven by a single site, a single device type, or one imaging workflow. [17] A particularly practical contribution is the work by Papachristou et al., where decision support was evaluated in primary care. If such systems are to improve melanoma detection in practice, the largest benefit may come at first contact, by supporting triage and identifying cases that require urgent dermatology referral. [16]

4.5 Variation in imaging conditions and the reliability of AI-based tools

Model performance depends on the data used for development and testing. If the training material is homogeneous and collected under ideal conditions (one camera type, one population, mainly standard dermoscopic images), performance may look excellent; difficulties often appear when conditions change. [15-17]

HAM10000 is one of the most frequently cited public dermoscopic image datasets and serves as an important reference for comparative research. It enables more standardised benchmarking, but it does not remove the fact that real-world clinical practice is more heterogeneous than any single dataset. [20] This is one reason why the literature increasingly includes datasets and studies based on whole-body imaging, where acquisition conditions, lesion counts, and patient context differ from classic dermoscopy.

4.6 Total body photography and 3D-TBP: where AI changes surveillance

TBP (including 3D-TBP) is especially valuable in patients with many naevi and increased skin cancer risk, because it allows creation of a skin "map" and supports follow-up over time. In practice, the main challenge of TBP is scale: very large numbers of lesions to review, comparisons between visits, and the risk of missing a new or changing lesion among many similar ones. [21-23]

In this context, datasets designed for automated methods are important. SLICE-3D is a dataset derived from 3D-TBP, containing hundreds of thousands of cropped lesion images, created to support development of algorithms for lesion detection and assessment in whole-body imaging. It addresses a practical need: models should be able to work with TBP-derived images, not only with selected, high-quality dermoscopic photographs. [21]

From a clinical implementation perspective, Winkler et al. evaluated an automated total body mapping algorithm for detecting clinically relevant melanocytic lesions. This is the type of tool that can reduce review time, structure the examination, and direct attention to lesions that warrant closer inspection. [22] From a clinician's viewpoint, it is not an "automatic melanoma detector", but a support tool that helps manage surveillance in patients with numerous lesions.

Ferreirinha et al. describe 3D-TBP as a method supporting melanoma diagnosis and show that current development is moving toward linking whole-body mapping with targeted analysis of suspicious lesions. At the same time, their work makes it clear that 3D-TBP is intended to be part of a diagnostic strategy rather than a replacement for clinical assessment and dermoscopy. [23]

4.7 How this can be combined into a single diagnostic workflow

Based on the cited studies, a clinically coherent workflow can be outlined:

- TBP/3D-TBP at baseline, to document the initial skin status and facilitate detection of new or changing lesions over time. [21-23]

- Mapping algorithms, to automatically highlight areas of interest and reduce the risk of missing relevant lesions. [22]

- Dermoscopy plus AI, to support assessment of selected lesions (for example new, evolving, or atypical) and to provide a second opinion in borderline cases. [10-17]

- Final clinical decision by the clinician, because responsibility remains with the physician; the tool should support decision-making rather than "issue verdicts". This is consistent with work emphasising human-AI collaboration and the effect of decision support on clinical choices. [13, 18]

4.8 Limitations that should be stated explicitly

These technologies are promising, but strong performance in one setting does not automatically translate to another. Differences in input data, patient populations, and imaging conditions can affect model performance; therefore, prospective and multicentre validation is essential. [15–17] A second key issue is the human-AI interaction: it is not enough for a model to perform well; its output must be presented in a way that supports sound clinical judgement rather than misleading clinicians or inappropriately inflating their confidence. [13, 18] Finally, TBP introduces an additional challenge: very large numbers of lesions and the need for meaningful selection of what is clinically relevant. [21,22]

4.9 Summary

Based on the cited bibliography, AI in dermoscopy can achieve high performance in image classification, with documented comparisons to dermatologists, and prospective studies suggest that these tools can be evaluated beyond controlled test conditions. [10-17] At the same time, the practical direction increasingly points toward combining dermoscopy with whole-body mapping (especially 3D-TBP) and tools that help clinicians manage surveillance in patients with many lesions. This includes moving from dataset development, through automated mapping, to targeted assessment of selected lesions. [20–23] The most reasonable conclusion is that this is not a replacement of clinicians by algorithms, but an attempt to improve detection and organise surveillance, provided that validation is robust and clinical use is well defined. [13, 15–19, 22–23]

5. Artificial Intelligence Applications in Psoriasis Severity Assessment and Therapeutic Monitoring: A Comprehensive Review

Artificial intelligence and machine learning have emerged as transformative technologies in dermatological practice, particularly demonstrating promising results in the automated assessment and monitoring of psoriasis severity [24].

Psoriasis, a chronic immune-mediated inflammatory skin disorder affecting approximately 2–4% of the global population, has traditionally relied on manual clinical scoring systems, most notably the Psoriasis Area and Severity Index (PASI), which suffers from significant inter-rater variability and time-consuming evaluation processes [24].

Recent advances in deep learning methodologies, particularly convolutional neural networks (CNNs), have demonstrated remarkable capability in predicting PASI scores with accuracy comparable to or exceeding that of experienced dermatologists [24].

A systematic review analyzing 30 studies encompassing 9,748 patients and 41,080 psoriasis images revealed that machine learning models trained on large, diverse datasets could achieve PASI predictions with mean absolute errors as low as 3.12, compared to the mean absolute error of 4.67 observed among certified dermatologists [24].

Advanced neural network architectures, including Cascade Mask Region-based CNNs with Swin Transformer backbones, have successfully addressed historical limitations in PASI scoring concerning inter-rater variability in assessing affected areas, with some models achieving automated PASI area score accuracy of 95.96% compared to dermatologists' performance of 76.43% [24].

Beyond image-based assessment, machine learning has been extended to evaluate body surface area (BSA) measurements through U-Net–based segmentation models achieving intraclass correlation coefficients of up to 0.96 [24].

Additionally, AI methods have been applied to reflectance confocal microscopy (RCM) image analysis, where the Multiscale Encoder–Decoder Network (MED-Net) demonstrated high performance in identifying diagnostically uninformative areas, achieving 82% sensitivity and 93% specificity [25].

The clinical applicability of AI-driven tools has progressed substantially from retrospective laboratory validation toward prospective real-world implementation, establishing a five-level framework for clinical translation [24].

While most published studies remain at Level 1 (single dataset training) or Level 2 (external validation), recent investigations have demonstrated prospective clinical utility at Levels 3 and 4, including portable devices for automated PASI assessment through comprehensive body imaging and digital image analysis that showed strong reproducibility and alignment with dermatologist-verified outcomes [24].

Notably, machine learning models incorporating non-image-based predictors such as serum biomarkers—including peripheral blood mononuclear cell profiling with flow cytometry to characterize memory T-helper cell subsets—have identified associations between specific cellular populations, particularly CCR6+ Th17 cells, and both psoriasis severity and the risk of progression to psoriatic arthritis [24].

The integration of visual overlay techniques that display detected uninformative areas on coregistered dermoscopic images provides clinicians with objective quality assurance metrics at both the time of image capture and interpretation, potentially reducing patient recalls for inadequate imaging and enabling performance assessment of imaging technicians [25].

Despite these promising developments, significant limitations persist in current AI applications for psoriasis assessment, including insufficient exploration of subjective disease manifestations beyond objective physical findings, limited consideration of lesion-specific location impacts on quality of life, and minimal integration of patient-reported outcomes and psychological measures into automated assessment algorithms [24].

Therapeutic monitoring through AI-enhanced platforms represents an emerging frontier, with remote monitoring tools demonstrating the capability for early detection of disease changes and timely therapeutic adjustments based on current disease status [24].

Machine learning models trained on multicenter prospective datasets have shown enhanced generalization capability for predicting therapy response, particularly through ensemble learning and gradient boosting methodologies that integrate clinical, demographic, and imaging data [24].

Nevertheless, the heterogeneity of published literature, the predominance of retrospective study designs, limited diversity of patient populations, and reliance on proprietary datasets continue to hinder systematic evaluation and large-scale clinical implementation [24].

Future advancement requires multicenter prospective studies with ethnically diverse datasets, development of explainable AI models, integration of comprehensive patient-centered outcome measures, and systematic evaluation of model architectures and training strategies. Artificial intelligence systems should ultimately function as supportive tools that augment—rather than replace—dermatological expertise while adequately addressing the multifaceted psychological and quality-of-life burden of psoriasis [24]

6. KEY TRENDS, LIMITATIONS, AND THREATS IN THE AUTOMATIC CLASSIFICATION OF SKIN LESIONS

In recent years, deep learning models have achieved promising results in skin lesion classification and segmentation task [26,27]. Despite the enormous benefits that AI brings, the use of artificial intelligence-based solutions in dermatology involves significant risks. The challenges associated with using artificial intelligence in dermatology stem primarily from the need to adapt algorithms to diverse skin types, limited data quality, variability of disease symptoms and lack of transparency in the operation of neural networks [27,29]. The shortage of images of skin diseases is a significant barrier to effective neural network training and the flow of data between their sources is insufficient. At the same time, the available materials lack uniform quality standards [27].

In the context of detecting skin disease such as melanoma, skin tone is a factor that must be annotated correctly but existing skin scales are often limited or inconsistently applied [28]. Most of the data sets used to train AI algorithms are heavily dominated by light skin types (Fitzpatrick I–III). Images of darker skin phototypes (Fitzpatrick IV–VI) are much less numerous, resulting in reduced model effectiveness for darker skin tones [27]. This data imbalance leads to marked differences in the sensitivity and specificity of algorithms between phototype groups. Improperly designed and tested AI systems may unknowingly exacerbate existing inequalities in healthcare [28]. Patients from underrepresented groups may receive less accurate diagnoses, which contradicts the principle of equal access to medical care. An additional limitation is the simplified treatment of skin phototype as a one-dimensional scale from light to dark, which does not adequately represent the heterogeneity of skin colors. Research findings indicate that algorithms may be less effective at recognizing skin lesions in individuals with certain skin tones, such as yellowish or olive tones, even if they belong to the same phototype [28]. Such subtle biases are difficult to identify unequivocally, but they can have significant clinical consequences.

Another aspect that warrants consideration is that the appearance of skin lesions in photographs is affected by differences in lighting, camera settings, camera quality, and device type, which can significantly influence the image features used by convolutional neural networks [27]. AI models often “learn” these technical differences related to image acquisition conditions rather than actual disease characteristics. As a result, the algorithm may only work well when photos are taken under conditions similar to those that are used to train the models. In the studies analyzed, neural networks were most often trained and evaluated on the same data set, but when implemented in a different center with a different population or using different diagnostic equipment, their performance drops dramatically [27]. This closed learning and testing scheme highlights a common limitation of machine learning methods, known as the generalization problem. In studies testing the generalization ability of models, neural networks often failed to achieve satisfactory results. The rare use of independent validation sets limits the ability to assess the clinical usefulness of an algorithm [27]. Machine learning algorithms largely reflect the nature of the data on which they are trained, so any biases present in the input image sets directly translate into their performance. The effects of these biases only become apparent when testing models on independent data sets.

From a legal perspective, issue that has not yet been fully resolved are difficulties in interpreting the algorithm's decision - the inability to explain how neural networks work [29]. Modern neural networks function like "black boxes" - they can generate a result but it remains difficult to identify the factors that drive their decision-making processes. It is important for doctors to understand the basis on which the system recommends a given diagnosis. If the algorithm cannot provide a clear justification, it is more difficult to trust its results. This problem also carries has legal and ethical implications, as it makes it difficult to assign responsibility for potential errors [29].

Artificial intelligence technologies have significant potential to support dermatological diagnostics, but without a systematic solution to problems related to data quality, algorithmic bias, and lack of model transparency, they may lead to increased healthcare inequalities and clinical risk [27-29].Recent work points to the need for improved data set representativeness, reliable external validation, and a critical approach to model decision explanation methods [27-29]. Future research should combine technical improvements with ethical and regulatory frameworks to enable the safe and equitable implementation of AI systems in clinical practice.

Possible solutions to the problems:

- increasing the diversity of data sets, especially in terms of skin phototypes and shades,

- documenting photography conditions to enable analysis of the impact of these variables,

- testing algorithms on data from different centers and populations,

- developing more reliable methods for explaining algorithmic decisions,

- introducing clear regulations on the accountability and auditing of AI systems.

7. SUMMARY

The reviewed literature demonstrates that artificial intelligence, particularly deep learning based on convolutional neural networks, has achieved substantial progress in dermatological image analysis, AI systems have shown strong performance in key tasks such as image classification, lesion segmentation and object detection, with several studies reporting diagnostic accuracy comparable to or exceeding that of board-certified dermatologists in controlled settings [7-12,19] . The integration of transfer learning has further enhanced model performance in scenarios with limited training data, making AI development feasible even for rare skin conditions [1].

Beyond retrospective performance evaluations, recent research increasingly emphasizes clinical implementation and human–AI collaboration [13-18]. Prospective and multicentre studies suggest that AI-based decision support can improve diagnostic accuracy, especially among less experienced clinicians, provided that algorithmic outputs are presented in a transparent and clinically meaningful manner [9,13-17] . The expansion of AI applications into total body photography and three-dimensional whole-body imaging represents an important step toward improving surveillance in high-risk patients with numerous lesions, where cognitive overload poses a significant challenge [21-23].

Despite these advances, significant limitations remain. Many AI models suffer from limited generalizability due to homogeneous training datasets, underrepresentation of darker skin phototypes and sensitivity to variations in image acquisition conditions [27,28]. The "black-box" nature of deep learning models also raises ethical, legal and practical concerns regarding accountability and trust in clinical decision-making [29]. Without careful attention to data quality, external validation and explainability, AI systems risk reinforcing existing healthcare disparities [27-29].

In conclusion, artificial intelligence holds considerable promise as a supportive tool in dermatology, with the potential to enhance diagnostic accuracy, streamline workflows and improve patient surveillance [10-17,21-23] . However, its successful and equitable integration into clinical practice requires robust multicentre validation, diverse and representative datasets, transparent model design and clearly defined regulatory frameworks [15-19,27-29]. AI should be viewed not as a replacement for dermatological expertise, but as an adjunct that augments clinical judgement and supports high-quality patient care [13,18].

Bibliography

1. De A, Sarda A, Gupta S, Das S. *Use of Artificial Intelligence in Dermatology*. Indian J Dermatol. 2020;65(5):352–357. doi: 10.4103/ijd.IJD_418_20. PMID: 33165383; PMCID: PMC7640800. https://pmc.ncbi.nlm.nih.gov/articles/PMC7640800/
2. https://www.isic-archive.com/
3. 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 Aug 14;5:180161. doi: 10.1038/sdata.2018.161. PMID: 30106392; PMCID: PMC6091241. https://pubmed.ncbi.nlm.nih.gov/30106392/
4. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, Gutman D, Halpern A, Helba B, Hofmann-Wellenhof R, Lallas A, Lapins J, Longo C, Malvehy J, Marchetti MA, Marghoob A, Menzies S, Oakley A, Paoli J, Puig S, Rinner C, Rosendahl C, Scope A, Sinz C, Soyer HP, Thomas L, Zalaudek I, Kittler H. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol. 2019 Jul;20(7):938-947. doi: 10.1016/S1470-2045(19)30333-X. Epub 2019 Jun 12. PMID: 31201137; PMCID: PMC8237239. https://pubmed.ncbi.nlm.nih.gov/31201137/
5. 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 Apr 9. doi: 10.1109/JBHI.2018.2824327. Epub ahead of print. PMID: 29993994. https://pubmed.ncbi.nlm.nih.gov/29993994/
6. Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S. Skin Cancer Classification With Deep Learning: A Systematic Review. Front Oncol. 2022 Jul 13;12:893972. doi: 10.3389/fonc.2022.893972. PMID: 35912265; PMCID: PMC9327733. https://pmc.ncbi.nlm.nih.gov/articles/PMC9327733/
7. Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017). https://doi.org/10.1038/nature21056
8. . Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, Thomas L, Enk A, Uhlmann L; Reader study level-I and level-II Groups; Alt C, Arenbergerova M, Bakos R, Baltzer A, Bertlich I, Blum A, Bokor-Billmann T, Bowling J, Braghiroli N, Braun R, Buder-Bakhaya K, Buhl T, Cabo H, Cabrijan L, Cevic N, Classen A, Deltgen D, Fink C, Georgieva I, Hakim-Meibodi LE, Hanner S, Hartmann F, Hartmann J, Haus G, Hoxha E, Karls R, Koga H, Kreusch J, Lallas A, Majenka P, Marghoob A, Massone C, Mekokishvili L, Mestel D, Meyer V, Neuberger A, Nielsen K, Oliviero M, Pampena R, Paoli J, Pawlik E, Rao B, Rendon A, Russo T, Sadek A, Samhaber K, Schneiderbauer R, Schweizer A, Toberer F, Trennheuser L, Vlahova L, Wald A, Winkler J, Wölbing P, Zalaudek I. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018 Aug 1;29(8):1836-1842. doi: 10.1093/annonc/mdy166. PMID: 29846502.
9. Tschandl, P., Rinner, C., Apalla, Z. *et al.* Human–computer collaboration for skin cancer recognition. *Nat Med* 26, 1229–1234 (2020). https://doi.org/10.1038/s41591-020-0942-0

10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. doi:10.1038/nature21056.
11. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166.
12. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer*. 2019;113:47-54. doi:10.1016/j.ejca.2019.04.001.
13. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. *Nat Med*. 2020;26:1229-1234. doi:10.1038/s41591-020-0942-0.
14. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. *Lancet Digit Health*. 2023;5(10):e679-e691. doi:10.1016/S2589-7500(23)00130-9.
15. Marchetti MA, Cowen EA, Kurtansky NR, Weber J, Dauscher M, DeFazio J, et al. Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study). *NPJ Digit Med*. 2023;6(1):127. doi:10.1038/s41746-023-00872-1.
16. Papachristou P, Söderholm M, Pallon J, Taloyan M, Polesie S, Paoli J, et al. Evaluation of an artificial intelligence-based decision support for the detection of cutaneous melanoma in primary care: a prospective real-life clinical trial. *Br J Dermatol*. 2024;191(1):125-133. doi:10.1093/bjd/ljae021.
17. Heinlein L, Maron RC, Hekler A, Haggenmüller S, Wies C, Utikal JS, et al. Prospective multicenter study using artificial intelligence to improve dermoscopic melanoma diagnosis in patient care. *Commun Med (Lond).* 2024;4(1):177. doi:10.1038/s43856-024-00598-5.
18. Krakowski I, Kim J, Cai ZR, Daneshjou R, Lapins J, Eriksson H, et al. Human-AI interaction in skin cancer diagnosis: a systematic review and meta-analysis. *NPJ Digit Med*. 2024;7(1):78. doi:10.1038/s41746-024-01031-w.
19. Salinas MP, Sepúlveda J, Hidalgo L, Peirano D, Morel M, Uribe P, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. *NPJ Digit Med*. 2024;7(1):125. doi:10.1038/s41746-024-01103-x.
20. 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. doi:10.1038/sdata.2018.161.
21. Kurtansky NR, D'Alessandro BM, Gillis MC, Betz-Stablein B, Cerminara SE, Garcia R, et al. The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D TBP for skin cancer detection. *Sci Data*. 2024;11(1):884. doi:10.1038/s41597-024-03743-w.
22. Winkler JK, Kommoss KS, Toberer F, Enk A, Maul LV, Navarini AA, et al. Performance of an automated total body mapping algorithm to detect melanocytic lesions of clinical relevance. *Eur J Cancer*. 2024;202:114026. doi:10.1016/j.ejca.2024.114026.
23. Ferreirinha A, Farricha V, João AL. Melanoma diagnosis with 3D total-body photography. *Actas Dermosifiliogr*. 2025;116(10):1116-1122. doi:10.1016/j.ad.2024.09.030.
24. McMullen EP, Al Naser YA, Maazi M, Grewal RS, Abdel Hafeez D, Folino TR, et al. Predicting psoriasis severity using machine learning: a systematic review. *Clin Exp Dermatol*. 2025;50:520–528. 2.
25. Kose K, Bozkurt A, Alessi-Fox C, Brooks DH, Dy JG, Rajadhyaksha M, et al. Utilizing machine learning for image quality assessment for reflectance confocal microscopy. *J Invest Dermatol*. 2020;140:1214–1222.
26. Izu-Belloso, R. M. (2025). *Generative Adversarial Networks in Dermatology: A Narrative ... PubMed Central (PMC).*
27. Fernandes, T. R. S., Teles, A. S., Fernandes, J. R. N., Lima, L. D. B., Sousa, D. L., & Soares, R. C. (2025). *External validation of AI models for skin diseases: A systematic review. IEEE Access.*

28. Xiang, A. (2023, October 3). *AI algorithms are biased against skin with yellow hues*. *WIRED*.
29. Murad, N. Y., Hasan, M. H., Azam, M. H., Yousuf, N., & Yalli, J. S. (2024). *Unraveling the Black Box: A review of explainable deep learning healthcare techniques*. *IEEE Access*.