International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A study on Detection of Skin Diseases from Digital Images

Mr. Shashidhara.B[1], Mr. Gururaj J P[2], Mr. Dadavali S P[3]

[1]Associate Professor in Computer Science

Abstract:
Skin diseases are very common, but early diagnosis is difficult because specialists are few, treatment is costly, and many diseases look alike. This work presents a survey of deep learning methods used for automatic skin disease detection from digital images. Twenty-three studies published were reviewed and compared on the basis of dataset, model architecture and reported results. The survey shows that Convolutional Neural Networks dominate this field and perform far better than general physicians, and almost equal to expert dermatologists, on well-defined tasks. However, it is observed that accuracy falls sharply on external data, ordinary clinical photographs and darker skin tones, and that high overall accuracy often hides poor detection of dangerous classes such as melanoma. Based on this analysis, our opinion is that such systems should be used only as fast, low-cost early screening and decision-support tools, and that future work must focus on balanced datasets, per-class reporting, explainability and external validation before clinical use.

Keywords: skin disease detection, deep learning, CNN, survey, transfer learning, explainable AI

1. INTRODUCTION

1.1 Overview

Skin is the largest organ of the human body and also the most exposed one, so it is easily affected by infections, allergies, and other disorders. Skin diseases are extremely common across the world. The Global Burden of Disease Study 2021 reported about 4.7 billion new cases of skin and subcutaneous diseases and around 2 billion existing cases in a single year, making them one of the top ten causes of disability worldwide. Because of this large burden, the World Health Organization has recognised skin diseases as a global public health priority. Some skin diseases are mild, like acne or fungal infection, but others are dangerous. Skin cancers are among the most commonly diagnosed cancers in the world, with more than 1.5 million new cases estimated in 2022. Early detection is very important, because most skin conditions can be treated easily if they are found at an early stage.

1.2 Background and Motivation

Normally, a skin disease is identified by a dermatologist who looks at the affected area, asks about symptoms, and sometimes takes a biopsy for laboratory testing. This process works well but has three problems. First, there are very few skin specialists compared to the number of patients, especially in rural and semi-urban areas. Second, visiting a specialist and getting lab tests takes time and money. Third, many skin diseases look very similar to each other, so even doctors can disagree. Studies show that general physicians and nurse practitioners are much less accurate than dermatologists when diagnosing skin

conditions from images, with top-1 accuracy around 0.44 for primary care physicians compared to 0.63 for dermatologists. This gap creates a strong need for a low-cost tool that can help in early screening.

1.3 Role of Deep Learning

Deep learning is a branch of artificial intelligence in which a computer model learns directly from data without being given fixed rules. For images, Convolutional Neural Networks (CNNs) are used because they automatically learn useful features such as colour, texture, shape, and the border of a lesion. Instead of a programmer writing rules like "if the patch is red and scaly, it may be psoriasis", the network learns these patterns itself from thousands of labelled images. Research in this area has produced very encouraging results. A systematic review found median diagnostic accuracies of about 94% for acne, 94% for rosacea, 93% for eczema, and 89% for psoriasis. For skin cancer, a meta-analysis reported an overall sensitivity of 87% and specificity of 77% for AI algorithms, which was better than generalists and comparable to expert dermatologists. These findings show that deep learning can act as a useful support system for doctors and patients.

1.4 Problem Statement

Even with such progress, a reliable and practical system for automatic skin disease detection is still not easy to build. Public datasets are limited in size, unbalanced across diseases, and often contain very few images of darker skin tones, which reduces fairness. Images taken by mobile phones vary in lighting, focus, distance, and background, so a model trained on clean clinical images may fail on real photographs. Deep learning models also behave like a "black box", giving a prediction without explaining the reason, which makes doctors hesitant to trust them. The problem addressed in this study is therefore to design and evaluate a deep learning based system that can classify common skin diseases from images with good accuracy, while handling these practical difficulties.

1.5 Scope of the Study

This study focuses on image-based detection of selected common skin diseases using deep learning. The system is intended only as a screening and decision support tool, not as a replacement for medical diagnosis. Final confirmation must always be done by a qualified doctor. The work is limited to publicly available datasets and standard CNN architectures, and does not include patient history, blood reports, or biopsy data.

2. LITERATURE SURVEY

2.1 Introduction

Automatic detection of skin diseases from images has been studied for more than a decade, but the field changed completely after 2017 when deep Convolutional Neural Networks (CNNs) were shown to match dermatologists on skin cancer images. This chapter reviews 23 important studies published between 2017 and 2025. For each study the dataset, the model used, the reported numbers and the main limitation are listed. After the table, the results of all these studies are compared and analysed, and the findings, observations and research gaps are presented.

2.2 Summary of Reviewed Literature

Sl.no.	Author(s) & Year	Dataset (size)	Model / Method	Reported Results	Limitation
1	Esteva et al., 2017, *Nature*	129,450 clinical images, 2,032 diseases in taxonomy, 757 training classes	Inception-v3, ImageNet pre-trained, fine-tuned	3-way accuracy 72.1 ± 0.9% vs dermatologists 65.56% and 66.0%; 9-way 55.4 ± 1.7%; binary AUC > 91%	Sensitivity/specificity given only in graphs; needs prospective clinical evaluation
2	Haenssle et al., 2018, *Annals of Oncology*	100 dermoscopic test images; 58 dermatologists	Inception-v4, binary melanoma vs nevus	CNN ROC AUC 0.86 vs dermatologists 0.79 ($p<0.01$); readers sens 86.6%, spec 71.3%; CNN spec 82.5% at matched sensitivity	Artificial 100-image reader setting, not real clinical practice
3	Brinker et al., 2019, *European Journal of Cancer*	Trained on 12,378 dermoscopic images; tested on 100 images vs 157 dermatologists	CNN, binary melanoma vs nevus	Dermatologists sens 74.1%, spec 60%; CNN spec 86.5% at same sensitivity, sens 87.5% at same specificity; beat 136 of 157 dermatologists	Very small image-only test set; accuracy and AUC not reported
4	Tschandl et al., 2019, *Lancet Oncology*	ISIC 2018: 10,015 train + 1,511 test, 7 classes; 139 algorithms vs 511 human readers	Challenge benchmark (many CNNs)	Algorithms 19.92 correct vs humans 17.91 (difference 2.01, $p<0.0001$); experts 18.78 vs top-3 algorithms 25.43; out-of-distribution drop 3.6% (algorithms) vs 11.4% (humans)	Performance decreases for out-of-distribution images
5	Liu et al., 2020, *Nature Medicine*	16,114 teledermatology cases, 26	Inception-v4 based DLS + clinical metadata	Top-1 0.71, top-3 0.93; on 963 cases top-1 0.66–	Private data; prospective clinical impact not assessed

		conditions (419 total)		0.67 vs dermatologists 0.63, PCPs 0.45, NPs 0.41	
6	Han et al., 2018, *J. Investigative Dermatology*	19,398 clinical images (Asan + MED-NODE + Atlas), 12 classes	ResNet-152 fine-tuned	Asan AUC: BCC 0.96, melanoma 0.96, SCC 0.83; Edinburgh AUC: SCC 0.91, melanoma 0.88; Hallym BCC sensitivity 87.1%	Needs wider range of ages and ethnicities
7	Wu et al. (AIDDA), 2020, *Ann. Translational Medicine*	4,740 images from 1,614 cases, 3 classes (psoriasis / eczema+AD / healthy)	EfficientNet-b4 with 7 auxiliary classifiers, 5-fold CV	Accuracy 95.80 ± 0.09%, sens 94.40 ± 0.12%, spec 97.20 ± 0.06%; AUC 0.987; psoriasis 89.46%, eczema+AD 92.57%	Eczema and atopic dermatitis had to be merged; 18% confusion with psoriasis; single centre
8	Tschandl et al., 2018, *Scientific Data* (HAM10000)	10,015 dermatoscopic images, 7 classes, >50% biopsy-confirmed	Dataset paper; Inception-v3 used only as image-type filter	Image-type classifier 98.68% top-1; covers >95% of pigmented lesions at the two sites	Only pigmented lesions; images differ from real-world user photos
9	Codella et al., 2019 (ISIC 2018 challenge)	Training 25,331 and test 8,238 images on the official portal	Benchmark for segmentation, attributes, classification	159 classification submissions; top segmentation still fails on >10% of images	Equal test scores do not mean equal generalisation
10	Choy et al., 2023, *npj Digital Medicine* (systematic review)	64 studies / 69 algorithms; median training set 2,555 images	CNN in 80% of algorithms; 88% used macroscopic images	Median accuracy: acne 94%, rosacea 94%, eczema 93%, psoriasis 89%, urticaria 81%; external accuracy falls (psoriasis 74%, skin infections 59%)	92% of studies at high risk of bias; only 19% report ethnicity or skin type; only 14% externally validated
11	Salinas et al., 2024, *npj*	53 comparative	CNN-based AI vs clinicians	AI sens 87.0% / spec 77.1% vs	AI advantage almost disappears

	Digital Medicine (meta-analysis)	studies, 19 pooled		clinicians 79.8% / 73.6%; vs generalists 92.5%/66.5% against 64.6%/72.8%; vs experts 86.3%/78.4% against 84.2%/74.4%	against expert dermatologists; 73.6% tested only internally
12	Groh et al., 2021, CVPR Workshops (Fitzpatrick 17k)	16,577 clinical images, 114 conditions, Fitzpatrick labels	Deep networks trained per skin-type partition	Models are "most accurate on skin types similar to those they were trained on"; skin type VI only 3.97% (635 images)	Darker skin under-represented; most public datasets have no skin-type labels
13	Daneshjou et al., 2022, *Science Advances* (DDI)	656 biopsy-proven images balanced across three skin-tone groups	Evaluation + fine-tuning of state-of-the-art dermatology models	ROC-AUC dropped 27–36% on diverse skin tones; all models worse on dark skin; fine-tuning closed the gap	DDI is too small to train a general model
14	Alipour et al., 2024, *Current Dermatology Reports*	Audit of 54 skin lesion datasets	Metadata review, not a classifier	Fitzpatrick type reported in only 3 of 54 datasets (5.56%); age in 35.19%; skin type VI is 1 image (0.07%) of PAD-UFES-20	Skin type diversity is either unmeasurable or inadequate
15	Bajwa et al., 2020, *Applied Sciences*	DermNet 21,844 images (23 / 622 classes); ISIC Archive 23,665 images, 7 classes	Ensemble of ResNet-152, DenseNet-161, SE-ResNeXt-101, NASNet + disease ontology	DermNet 23-class top-1 77.53%, ontology 79.94%, top-5 93.87%, AUC 97.60%; 622-class top-1 66.74%; ISIC 7-class 93.06%, AUC 99.23% but	High overall accuracy hides weak classes (cellulitis F1 51.92%)

				melanoma recall only 66.05%	
16	Ding et al., 2023, *Digital Health* (HI-MViT)	HAM10000 10,015 (7 classes), ISIC-2017 2,750 (3 classes), PH² 200	Lightweight modified MobileViT (4.9M parameters)	ISIC-2018 accuracy 0.932 ± 0.002, F1 0.931, AUC 0.977; PH² transfer 0.914; MobileNet 0.867, ShuffleNet 0.869, plain ViT only 0.742 (86.2M params)	Still validated mainly on curated public benchmarks
17	Paraddy & Virupakshappa, 2024, *J. Imaging Informatics in Medicine*	HAM10000 10,015; ISBI 2016 2,558; PH2 200; Skin Cancer ISIC 2,357	Hybrid CNN + Swin Transformer (Swinformer-Net + MD-CNNFormer)	Overall accuracy 98.72%; HAM10000 accuracy 98.94%, precision 99.23%, recall 98.5%, F1 98.18%, AUC 0.985	Very high numbers on the same public sets; no external clinical validation
18	Hossain et al., 2023, *Diagnostics*	ISIC 2018 + HAM10000	Max-voting ensemble of 10 pre-trained CNNs	Individual models 77.20%–91.90%; ensemble accuracy 93.18%, AUC 0.9320	Dataset sizes, class counts and F1 not reported
19	Manole et al., 2024, *Bioengineering*	8,222 images (own + ISIC 2019) → 12,864 after augmentation, 6 classes	Custom EfficientNetB3 with batch-norm and dropout 0.45	4-class accuracy 95.4%, melanoma precision 1.00 / recall 0.99, AUC 0.98–1.0; 6-class accuracy 88.8%, SCC F1 only 0.61	Limited demographic diversity; small dataset risks overfitting; poor interpretability
20	Halder et al., 2025, *Scientific Reports*	HAM10000 10,015 rebalanced to 42,000 (6,000 per class); DermaMNIST 10,015	Fuzzy rank-based ensemble of Xception + InceptionResNet V2 + MobileNetV2	HAM10000 ensemble 95.14% vs Xception 94.79%, MobileNetV2 94.30%, InceptionResNet V2 79.17%; DermaMNIST 78.25%	Balanced set created by heavily augmenting as few as 115 dermatofibroma images; same-source evaluation only

21	Srinivasu et al., 2021, *Sensors*	HAM10000, 7,224 train / 1,255 validation, 7 classes	MobileNet V2 + LSTM	Accuracy 85.34%, sens 88.24%, spec 92.00%, MCC 86.00% vs MobileNetV2 84.00%, VGG19 81.00%, plain CNN 80.00%	Class imbalance needed augmentation; model size and latency not reported
22	Fraiwan & Faouri, 2022, *Sensors*	HAM10000 10,015, 7 classes (NV 6,705 … DF 115)	13 transfer-learning models, 10 runs each	Best accuracy ResNet101 76.7%; DenseNet201 75.8% (F1 64.8%); SqueezeNet 71.7% (F1 51.2%); melanoma precision only 43.1% at recall 71%	Severe class imbalance inflates specificity while melanoma precision stays low
23	Velasco et al., 2019, *IJATCSE*	3,406 dermatologist-validated clinical photographs, 7 classes	MobileNet transfer learning in an Android app	Accuracy 94.4% with oversampling + augmentation (91.8% oversampling only, 84.28% undersampling); weight size 16.823 MB, load time 4.838 s	Psoriasis still confused with acne and pityriasis rosea; only 7 conditions

2.3 Datasets Used in the Literature

Dataset	Images	Classes
Esteva et al. corpus	129,450 (3,374 dermoscopic)	2,032 in taxonomy, 757 trained
HAM10000	10,015	7 (NV 6,705, MEL 1,113, BKL 1,099, BCC 514, AKIEC 327, VASC 142, DF 115)
ISIC 2018 challenge	25,331 train / 8,238 test	7
ISIC 2016 / 2017	900 / 2,000–10,015	2–3
SIIM-ISIC 2020	33,126 train / 10,982 test	binary

DermNet	21,844	23 super-classes / 622 sub-classes
Fitzpatrick 17k	16,577 (16,012 with known skin type)	114 conditions, 6 skin types
Diverse Dermatology Images (DDI)	656	3 skin-tone groups
PAD-UFES-20	2,299	skin type VI only 1 image
PH²	200 (160 nevi, 40 melanoma)	3
BCN20000	19,424	—
Asan (full)	120,780	—
XiangyaDerm	107,565	—
Derm101	107,656	—
MoleMap	102,451	—
DermaMNIST	10,015 at 3 × 28 × 28	7

2.4 Findings

1. Deep learning, mainly CNNs, is the dominant approach used in 80% of algorithms in a 64-study review and has repeatedly reached dermatologist-level accuracy on narrow, well-defined tasks.
2. AI clearly outperforms general physicians and nurse practitioners (pooled sensitivity 92.5% vs 64.6%) but is only comparable to expert dermatologists (86.3% vs 84.2%).
3. Accuracy depends heavily on task design: binary and few-class tasks report 93–99%, while fine-grained tasks with hundreds of classes fall to 55–67%.
4. Image modality matters: dermoscopic benchmarks give much higher scores than ordinary clinical photographs taken with variable lighting and angle.
5. Overall accuracy is a misleading metric under class imbalance. Melanoma recall of 66.05% and precision of 43.1% were reported inside models advertised as 93% accurate.
6. Very few studies validate externally (14%), and accuracy drops by 15–29 percentage points when they do.
7. Datasets are demographically narrow. Skin type is almost never recorded, dark skin is severely under-represented, and models lose 27–36% AUC on diverse skin tones.
8. Ensembles and hybrid CNN-Transformer models consistently beat single networks (e.g. 93.18% ensemble vs 77.20–91.90% individually; 95.14% fuzzy ensemble vs 79.17–94.79% individually).
9. Lightweight architectures are viable for mobile use, reaching 93–94% accuracy at under 5M parameters or about 17 MB, which supports a phone-based screening application.
10. Reporting quality is weak: 92% of studies in one review were judged at high risk of bias, and many papers omit sensitivity, specificity, F1 or dataset composition.

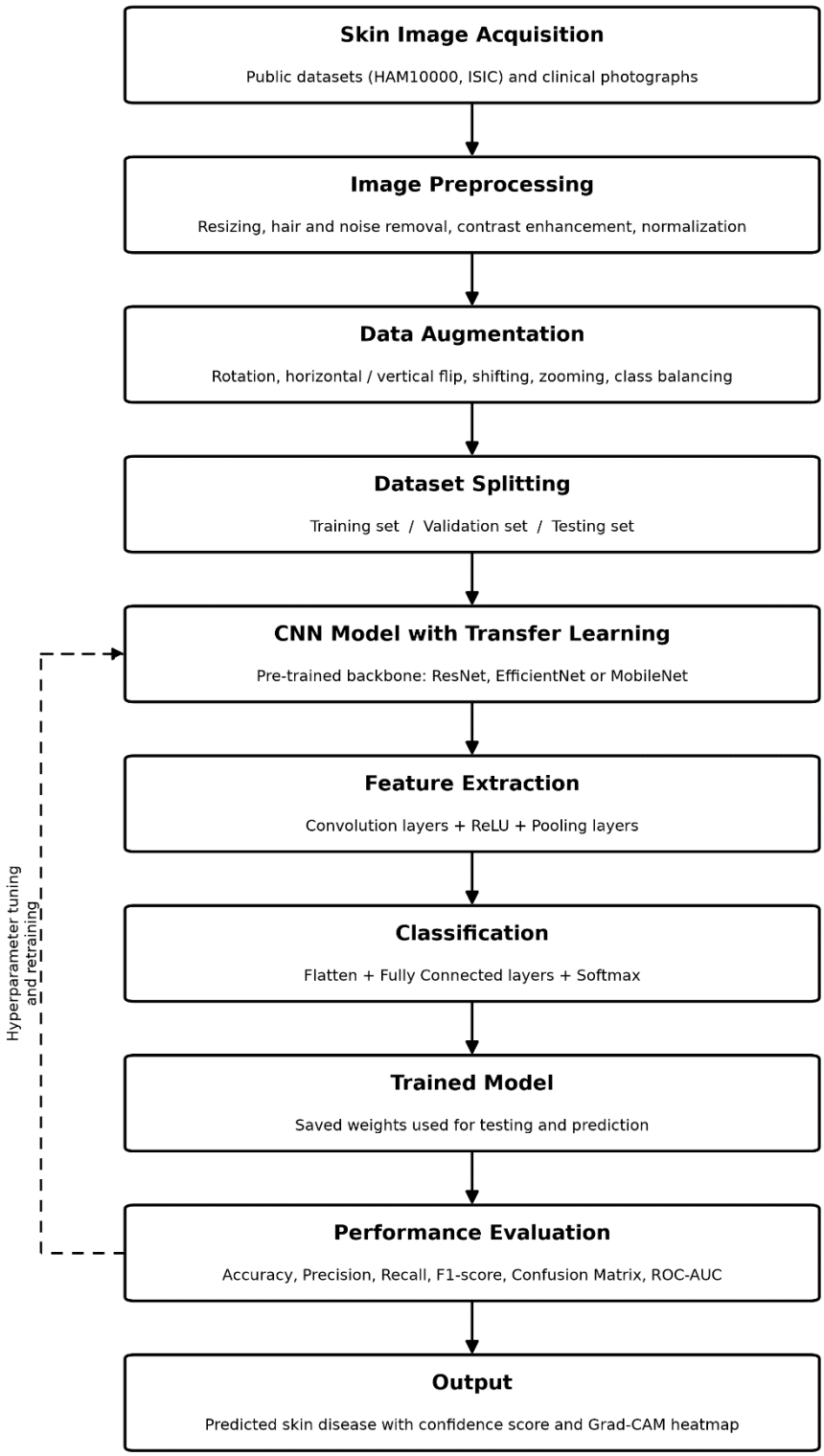
2.5 Observations

1. Papers claiming 98–99% accuracy and rigorous multi-model studies topping out at 76.7% coexist on the *same* dataset. A student project should therefore treat any accuracy above about 95% on HAM10000 with caution and check the split, the augmentation and whether the test set was truly held out.

2. Merged classes hide real difficulty. AIDDA had to combine eczema and atopic dermatitis because they cannot be separated by appearance alone. Any class list must be clinically defensible, not just convenient.
3. Augmentation is not new information. Expanding 115 real dermatofibroma images into 6,000 synthetic ones raises balance but not true diversity, and can inflate validation scores.
4. Top-k reporting is more honest than top-1 for fine-grained tasks: 622-class top-1 was 66.74% but top-5 was 86.26%, which is still clinically useful as a shortlist.
5. Explainability is repeatedly listed as a barrier. Without Grad-CAM style visual evidence, clinicians are unlikely to trust the output.
6. Fairness must be designed in from the start, since fine-tuning on diverse images has been shown to remove the light/dark performance gap.
7. Because AI mainly beats non-specialists, the highest practical value lies in primary care and rural screening, not in replacing dermatologists.
8. Reader-study designs using 100 curated images cannot predict real clinical performance; prospective evaluation is still largely missing from the literature.

3. PROPOSED SYSTEM

Fig. 3.1 Block Diagram of the Proposed System



The proposed system takes a photograph of the affected skin and predicts the disease using a deep learning model. The image is first cleaned and resized, and then a Convolutional Neural Network extracts features like colour, texture and shape to classify it. Transfer learning and data augmentation are used so that the

model works well even with a limited number of medical images. The system gives the disease name with a confidence score within seconds and acts as an early screening tool to support doctors, not replace them.

1. Skin Image Acquisition images are taken from public datasets such as HAM10000 (10,015 dermatoscopic images, 7 classes) and the ISIC archive, or captured directly as clinical photographs.
2. Image Preprocessing all images are resized to a fixed size, unwanted hair and noise are removed, contrast is improved, and pixel values are normalised.
3. Data Augmentation rotation, flipping, shifting and zooming increase the number of images and reduce class imbalance, which is severe in HAM10000 (6,705 nevus images against only 115 dermatofibroma).
4. Dataset Splitting the data is divided into training, validation and testing sets so that testing images are never seen during training.
5. CNN Model with Transfer Learning a pre-trained backbone (ResNet, EfficientNet or MobileNet) is fine-tuned instead of training from zero, which works better on small medical datasets.
6. Feature Extraction convolution, ReLU activation and pooling layers automatically learn colour, texture, shape and border features of the lesion.
7. Classification flatten, fully connected layers and a softmax layer give the probability of each disease class.
8. Trained Model the learned weights are saved and reused for prediction on new images.
9. Performance Evaluation accuracy, precision, recall, F1-score, confusion matrix and ROC-AUC are calculated. Per-class metrics are important because overall accuracy can hide poor melanoma detection.
10. Output — the predicted disease name with a confidence score, plus a Grad-CAM heatmap showing which region influenced the decision.

CONCLUSION

This work studied the detection of skin diseases from images using deep learning and showed that Convolutional Neural Networks can learn disease patterns directly from data without any manually written rules. The literature clearly proves that such models reach dermatologist-level accuracy on well-defined tasks, and are far better than general physicians in early screening. However, the accuracy drops when the model is tested on new hospitals, ordinary phone photographs, a larger number of classes, or darker skin tones, which shows that high reported accuracy alone is not enough. Class imbalance also hides poor detection of dangerous conditions like melanoma, so per-class metrics must always be reported. The proposed system therefore uses transfer learning, data augmentation and Grad-CAM based explanation to give fast, low-cost and reasonably reliable results. In conclusion, deep learning is a practical support tool for early skin disease screening, especially in rural areas, but the final diagnosis must always be confirmed by a qualified dermatologist.

REFERENCES:

1. Esteva A. et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118.
2. Haenssle H.A. et al. (2018). Man against machine. *Annals of Oncology* 29(8):1836.
3. Brinker T.J. et al. (2019). Deep learning outperformed 136 of 157 dermatologists. *European Journal of Cancer*.

4. Tschandl P. et al. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms. *Lancet Oncology* 20(7):938–947.
5. Liu Y. et al. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine* 26:900–908.
6. Han S.S. et al. (2018). Classification of clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*.
7. Wu H. et al. (2020). A deep learning, image based approach for automated diagnosis for inflammatory skin diseases (AIDDA). *Annals of Translational Medicine*.
8. Tschandl P., Rosendahl C., Kittler H. (2018). The HAM10000 dataset. *Scientific Data* 5:180161.
9. Codella N. et al. (2019). Skin lesion analysis toward melanoma detection 2018: ISIC challenge.
10. Choy S.P. et al. (2023). Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. *npj Digital Medicine*.
11. Salinas M.P. et al. (2024). A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. *npj Digital Medicine* 7:125
12. Groh M. et al. (2021). Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset. CVPR Workshops.
13. Daneshjou R. et al. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*.
14. Alipour N., Burke T., Courtney J. (2024). Skin type diversity in skin lesion datasets: a review. *Current Dermatology Reports*.
15. Bajwa M.N. et al. (2020). Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences* 10(7):2488.
16. Ding J. et al. (2023). HI-MViT: a lightweight model for explainable skin disease classification. *Digital Health* 9.
17. Paraddy V., Virupakshappa (2024). CSwinformer for skin lesion segmentation and classification. *Journal of Imaging Informatics in Medicine* 38(3):1755–1775.
18. Hossain M.M. et al. (2023). The segmentation of skin lesions and their classification with a max-voting ensemble. *Diagnostics* 14(1):89.
19. Manole A. et al. (2024). An interoperable deep learning approach for skin lesion classification. *Bioengineering* 11(8):810.
20. Halder A. et al. (2025). Fuzzy rank-based deep ensemble for skin cancer classification. *Scientific Reports* 15:6268.
21. Srinivasu P.N. et al. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* 21(8):2852.
22. Fraiwan M., Faouri E. (2022). On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors* 22(13):4963.
23. Velasco J. et al. (2019). A smartphone-based skin disease classification using MobileNet CNN. *IJATCSE* 8(5):2632–2637.