

Deep Learning for Skin Diseases Classification from Dermoscopic Images: A Comprehensive Survey

Chahinez GUERMAT
Department of informatics
University of ALMERIA
Carretera Sacramento, s/n 04120
La Cañada de San Urbano
Almería, Spain
gc090@inlumine.ual.es

Larbi BOUBCHIR
LIASD Laboratory
University of Paris 8
2 rue de la Liberté, 93526
Saint-Denis, France
larbi.boubchir@univ-paris8.fr

V. González-Ruiz
Department of informatics
University of ALMERIA
Carretera Sacramento, s/n 04120
La Cañada de San Urbano
Almería, Spain
vrui@ual.es

Abstract—Skin diseases encompass a broad spectrum of conditions, ranging from benign dermatological disorders to aggressive forms of skin cancer, for which early and accurate diagnosis is critical. In recent years, deep learning-based approaches have significantly advanced automated skin disease classification, particularly through the analysis of dermoscopic images. This paper presents a comprehensive and structured survey of recent deep learning techniques applied to skin disease classification. We review convolutional neural network (CNN) based models, hybrid architectures, and transformer-based approaches, highlighting their methodological principles and comparative performance. In addition, we analyze the main publicly available dermatological datasets commonly used for model development and evaluation, including ISIC Archive, HAM10000, PH², and DermNet, with particular attention to their strengths and limitations. Beyond summarizing existing methods, this survey discusses persistent challenges such as data quality, class imbalance, reproducibility, and clinical relevance. By synthesizing recent advances and identifying open research issues, this work aims to provide a clear reference framework to support the development of more robust and clinically reliable automated diagnostic systems in dermatology.

Index Terms—Skin Diseases, Deep Learning, Dermoscopic Images, Clinical Images, Convolutional Neural Network (CNN), Transformer.

I. INTRODUCTION

Skin diseases constitute a heterogeneous group of medical conditions, ranging from common inflammatory disorders to life-threatening skin cancers. In many cases, patient prognosis strongly depends on early and accurate diagnosis. In routine clinical practice, diagnosis is primarily based on visual examination, often supported by dermoscopy [1]. Although dermoscopy improves diagnostic accuracy, its effectiveness remains highly dependent on clinician expertise and subjective interpretation, which can lead to inter-observer variability [2]. These limitations have motivated increasing interest in automated computer-aided diagnostic systems to support dermatological decision-making.

The rapid progress of deep learning has opened new perspectives for image-based medical diagnosis, particularly in dermatology [2]. Among the proposed techniques, convolutional neural networks (CNNs) have been extensively studied due to their ability to automatically learn hierarchical visual

representations from dermoscopic images [3]. Early CNN-based approaches demonstrated promising performance; however, their effectiveness is often constrained by challenges such as limited labeled data, class imbalance, and reduced generalization across datasets [4].

To address these limitations, several studies have explored hybrid architectures that combine CNNs with additional learning components, including classical machine learning classifiers, segmentation modules, or data rebalancing strategies. More recently, transformer-based models have attracted growing attention in skin disease classification. By leveraging self-attention mechanisms, these architectures are capable of modeling long-range dependencies and capturing global contextual information within lesion images, offering an alternative to purely convolutional representations [4].

The availability of publicly accessible dermatological datasets has played a crucial role in advancing research in this field. Benchmark datasets such as ISIC Archive, HAM10000, and PH² have enabled standardized evaluation of deep learning models under controlled dermoscopic conditions. In contrast, DermNet has been widely used to assess model generalization on real-world clinical images. Despite substantial progress, persistent challenges remain, including dataset quality, class imbalance, annotation reliability, and limited reproducibility across studies.

In this context, the objective of this paper is to provide a structured and up-to-date survey of deep learning approaches for skin disease classification. By systematically reviewing CNN-based, hybrid, and transformer-based methods, and by analyzing the datasets on which these methods are evaluated, this survey aims to clarify current research trends, highlight methodological limitations, and identify open challenges that must be addressed to improve the clinical reliability of automated dermatological diagnostic systems.

The remainder of this paper is organized as follows. Section II presents the main publicly available dermatological datasets commonly used for skin disease classification. Section III reviews deep learning approaches, including CNN-based, hybrid, and transformer-based methods. Section IV discusses the strengths, limitations, and open challenges identified in the

literature, and Section V concludes the paper.

II. DATASETS

Public dermatological datasets are essential for the design and assessment of deep learning approaches in skin disease analysis. In this review, we consider four representative datasets that are summarized in Table I. ISIC Archive, HAM10000, and PH² are commonly used benchmark datasets in dermatological image analysis, whereas DermNet, although highly recognized, is better regarded as a reference data set for evaluating model generalization under real-world clinical conditions.

The **ISIC Archive** constitutes a large-scale collection of dermoscopic images acquired from multiple clinical institutions and released across several editions [5]. It has supported both segmentation and classification tasks, with a clear evolution from segmentation-oriented studies in earlier editions to classification-focused research in later releases, particularly melanoma detection. Representative examples of common skin lesion classes from the ISIC dataset are shown in Figure 1. Despite challenges related to class imbalance, image quality variability, and annotation noise, ISIC remains a cornerstone benchmark in dermatological image analysis [6].

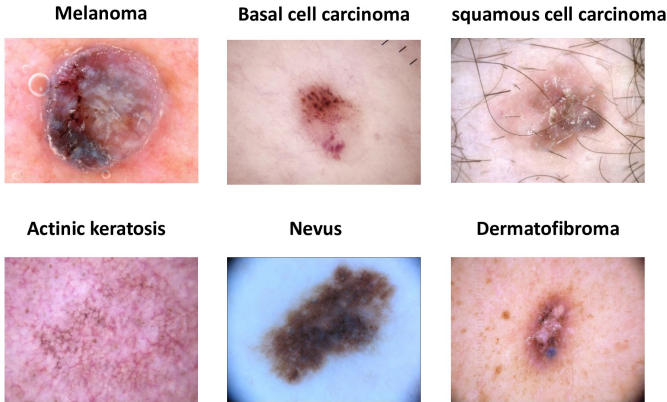


Fig. 1: Representative examples of common skin lesion classes from the ISIC dataset.

HAM10000 complements ISIC by providing a standardized dermoscopic dataset with expert annotations, making it suitable for multi-class skin lesion classification under controlled acquisition conditions. However, issues such as intra-class variability, inter-class similarity, and class imbalance persist and have motivated the adoption of advanced learning strategies to improve classification performance [7].

PH² is a smaller yet carefully annotated dermoscopic dataset that is primarily used for validation and comparative studies. Its limited size restricts its use for large-scale training, but it remains valuable for benchmarking and for evaluating the robustness and generalization of classification and segmentation models in controlled experimental settings [8].

DermNet mainly consists of clinical photographs captured under real-world conditions. Unlike the aforementioned dermoscopic datasets. It introduces substantial variability in illu-

mination, background, and lesion appearance. As a result, it is commonly employed to evaluate the generalization capability of deep learning models beyond the controlled acquisition environments. Thereby bridging the gap between research oriented datasets and real clinical practice [9].

TABLE I: Overview of publicly available dermatological datasets used for skin disease classification.

Dataset	Type of images	Number of images	Number of classes	Country
ISIC(2016-2020)	Dermoscopic	76000	2-8	International
HAM10000	Dermoscopic	10015	7	Austria
DermNet	Clinical	> 20000	23	New Zealand
PH ²	Dermoscopic	200	3	Portugal

Based on the datasets presented above, The next section provides a comprehensive review of existing approaches, with a particular emphasis on CNN-based, hybrid, and transformer-based methods developed to address these challenges.

III. LITERATURE SURVEY

This literature survey reviews state-of-the-art deep learning approaches for skin disease classification, with a particular focus on CNN-based and hybrid models, followed by transformer architectures.

A. CNN-based Approaches

Convolutional Neural Networks (CNNs) are special kind of neural networks that preserve the spatial relationships in data with very few connections between layers. They are able to extract meaningful representations from input data, which are suitable for image-oriented problems. A CNN consists of multiple layers of convolutions and activations, with pooling layers interspersed between different convolution layers. Figure 2 shows the Convolution and the Pooling operations.

It is trained via backpropagation similar with the standard neural networks. Additionally, a CNN typically includes fully connected layers at the end of the architecture to produce the output [10].

CNNs have been extensively explored for automated skin disease classification due to their strong capability to learn hierarchical visual representations from the dermoscopic images. Early studies primarily focused on designing custom CNN architectures or combining deep features with traditional machine learning classifiers to improve diagnostic performance.

Several works have proposed customized CNN models trained on curated dermatological image datasets. For example, Saifan and Jubair [11] introduced a CNN-based framework for classifying six common skin diseases using a balanced dataset collected from multiple online sources. In addition to model development, their study emphasized practical applicability through the implementation of a mobile diagnostic application. While such approaches demonstrate the feasibility of CNN-based diagnosis, their reliance on non-standardized datasets may limit generalization.

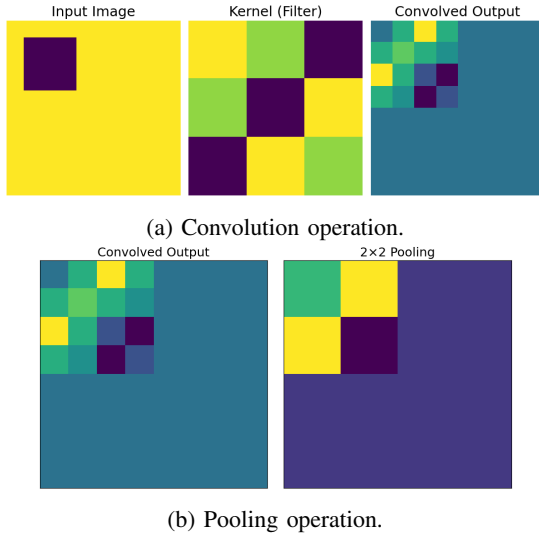


Fig. 2: Overview of the CNN components

Transfer learning has become a dominant strategy to mitigate data scarcity in medical imaging. Pre-trained architectures such as VGG-16, ResNet, DenseNet, and EfficientNet have been fine-tuned for skin disease classification with notable success. Saranya et al. [12] employed a VGG-16-based model enhanced through fine-tuning and preprocessing techniques, reporting high classification accuracy. However, the absence of detailed dataset description and clinically relevant evaluation metrics raises concerns regarding reproducibility and clinical reliability.

Comparative studies have also examined CNNs against traditional classifiers. Kushartanto et al. [13] evaluated CNN-based models and support vector machines (SVMs) using features extracted from convolutional layers. Their results indicated that CNNs consistently outperform SVMs in terms of accuracy, confidence scores, and computational efficiency. These findings reinforce the advantage of end-to-end deep learning models over handcrafted or hybrid feature-based classifiers.

The effectiveness of lightweight and deep CNN architectures has also been investigated. Harbola et al. [14] compared ResNet50 and MobileNetV2, demonstrating that lightweight architectures can achieve competitive performance with reduced computational cost. Nonetheless, the limited size of the training dataset restricts the robustness of the reported results, particularly for deeper models.

More advanced transfer learning pipelines have integrated preprocessing, segmentation, and class rebalancing to improve classification performance. Gururaj et al. [3] proposed a comprehensive framework combining lesion segmentation, sampling strategies, and fine-tuned DenseNet and ResNet models on the HAM10000 dataset. Similarly, Ali et al. [15] explored multiple EfficientNet variants, achieving strong performance through architecture-specific fine-tuning and data augmentation. These studies highlight the importance of tailored prepro-

cessing and model adaptation in dermatological applications.

Recent works have moved toward hybrid CNN architectures that combine features from multiple deep networks. Hasan et al. [16] proposed a multi-branch CNN integrating ResNet, Xception, and DenseNet components, achieving high AUC scores across multiple ISIC benchmarks. Other hybrid approaches combine CNN feature extractors (VGG) with classical classifiers such as random forests [17] or SVMs [18], [19]. While these models demonstrate promising performance, many rely on private or limited datasets, which restricts reproducibility and raises concerns regarding real-world deployment.

Despite significant progress, CNN-based approaches continue to face notable challenges. Many studies depend on small or non-public datasets, suffer from class imbalance, or evaluate performance using limited metrics. Addressing these issues through standardized benchmarks, transparent evaluation protocols, and diverse datasets remains essential for developing clinically reliable skin disease classification systems. An overview of CNN-based and hybrid approaches is provided in Table II.

B. Transformer-based Approaches

Transformer-based architectures have recently gained significant attention in dermatological image analysis due to their ability to model long-range dependencies and capture both local and global contextual information. Unlike conventional CNNs, vision transformers first decompose input images into a sequence of fixed-size patches, which are then processed through self-attention mechanisms. This patch-based representation, combined with self-attention, enables a more holistic modeling of complex lesion patterns by explicitly learning global relationships between different image regions [20]. The ViT architecture proposed by Dosovitskiy et al. [21] is presented in Figure 3.

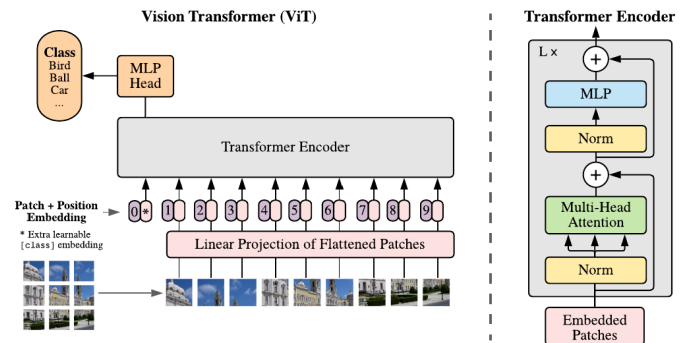


Fig. 3: Vision Transformer (ViT) Architecture from [21].

Several studies have explored the application of Vision Transformers (ViT) and their variants for skin disease classification. Saha et al. [22] proposed a transformer-driven framework that integrates ViT and Swin Transformer models with YOLOv8-based lesion segmentation. By isolating lesion regions prior to classification, their approach demonstrated

TABLE II: Related work on CNN-based and hybrid models for skin disease classification

Reference	Year	Dataset	Model	segmentation	augmentation	Metrics	Result
[11]	2022	collected	CNN	-	-	Acc	81.75%
[15]	2022	HAM10000	EfficientNet(B0-B7)	-	Yes	F1-score Top1-acc	87.91% 87.91%
[16]	2022	ISIC 2016–2018	ResNet Xception DensNet	DSNET	Yes	Aucc	0.96% 0.95% 0.97%
[3]	2023	HAM10000	ResNet50 DensNet169	Auto-encoder	Sampling	Acc	95.29% 97.44%
[12]	2024	-	VGG-16 (CNN)	-	-	Acc	99.05%
[13]	2024	Online-data	CNN + SVM	-	Yes	Acc	83.8%
[17]	2024	Collected	VGG-16 + RF	-	Yes	Acc	96.08%
[18]	2024	Private	CNN + SVM	-	Yes	F1-Score	75.28%
[19]	2024	Collected	CNN + SVM	-	Yes	Acc	94.12%
[14]	2025	From kaggle	ResNet50 MobileNetV2	-	-	Acc	95.29%

superior performance compared to conventional CNN architectures on the ISIC 2019 dataset. This highlights the importance of combining accurate lesion localization with transformer-based feature learning.

The effectiveness of hierarchical transformer architectures has also been demonstrated in multi-class classification tasks. Ahmad et al. [23] employed the Swin Transformer to classify a large number of dermatological conditions, leveraging extensive data augmentation to improve generalization. Their results confirm the suitability of window-based attention mechanisms for capturing fine-grained lesion details while maintaining global context.

More comprehensive transformer-based pipelines have recently emerged. Mohan et al. [4] investigated the integration of ViT, Swin Transformer, and DINOv2 for large-scale skin disease classification using a combined ISIC and DermNet dataset. Their results showed that self-supervised transformer representations, particularly DINOv2, outperform both CNN-based models and supervised transformer variants. The inclusion of explainable AI techniques further strengthens the clinical relevance of their approach.

Transformers have also been combined with advanced data augmentation strategies to address dataset imbalance. Wang et al. [24] introduced a two-stage framework in which synthetic melanoma images generated by a StyleGAN-based model were used to enhance training data. A modified Vision Transformer with batch-level self-attention was subsequently applied, leading to improved melanoma classification perfor-

mance. This work demonstrates the complementary benefits of generative modeling and transformer-based classification. Several studies have proposed customized transformer architectures tailored for dermoscopic image analysis. Yang [25] introduced the ViTfSCD model, which incorporates patch-based tokenization, multi-head self-attention, and transfer learning to achieve strong performance on the HAM10000 dataset. Similarly, Xin et al. [26] proposed the SkinTrans framework, introducing multi-scale patch partitioning and contrastive learning to enhance feature discrimination. While these approaches achieve high accuracy, their computational complexity and limited external validation remain key challenges for clinical deployment. Beyond pure transformer models, recent research has increasingly focused on hybrid transformer-based architectures. Sirawich et al. [27] proposed a multimodal framework combining a vision transformer for image encoding with CNN-based clinical metadata processing. Although effective for binary classification, the model's scope remains limited for real-world multi-class diagnostic scenarios. Other hybrid designs integrate transformers with CNN backbones to exploit both local feature extraction and global attention. Somaiya and Ali [28] proposed a hybrid SkinViT model incorporating Outlooker blocks, transformer encoders, and an SVM classifier. Hao et al. [29] introduced a ConvNeXt–Swin Transformer hybrid architecture with attentional feature fusion, demonstrating the benefits of combining convolutional and transformer representations. Despite promising results, many hybrid approaches rely on private datasets or binary classifi-

cation tasks, which limits reproducibility and clinical generalization. Overall, transformer-based and hybrid transformer architectures represent a rapidly evolving research direction in skin disease classification. While they often outperform CNN-based models, their high computational cost, dependence on large datasets, and limited interpretability remain open challenges that must be addressed for widespread clinical adoption. A summary of transformer approaches reviewed in this study is reported in Table III.

IV. DISCUSSION

The literature survey reveals a clear evolution of deep learning approaches for skin disease classification, progressing from conventional CNN-based architectures toward hybrid models and more recently transformer-based frameworks. CNN-based methods remain dominant in the literature, particularly when applied to dermoscopic benchmark datasets such as ISIC and HAM10000, where they demonstrate strong performance in capturing local lesion patterns. Hybrid approaches, combining CNNs with classical classifiers or complementary modules, are often adopted to improve robustness, especially in scenarios involving limited data or class imbalance.

Transformer-based models have recently gained increasing attention, as highlighted in the surveyed studies, due to their ability to model global contextual information through self-attention mechanisms. Several works report competitive or superior performance compared to CNNs, particularly when large-scale datasets, extensive data augmentation, or self-supervised pretraining strategies are employed. However, the literature also indicates that these gains often come at the cost of higher computational complexity and increased data requirements, which may limit their practical applicability in real clinical environments.

Across both CNN and transformer approaches, persistent challenges are observed. Dataset limitations, including class imbalance, annotation noise, and insufficient reporting of data sources in some studies, negatively impact reproducibility and fair comparison of results. In addition, many works focus primarily on accuracy-based metrics, with limited consideration of clinically meaningful evaluation criteria or external validation. Although some recent studies incorporate explainable AI techniques, interpretability remains insufficiently addressed in the majority of reviewed approaches.

Future research should therefore aim to develop more data-efficient and generalizable models that can operate reliably across heterogeneous datasets. The adoption of standardized evaluation protocols, transparent dataset reporting, and systematic integration of explainability mechanisms will be crucial for improving clinical trust. Furthermore, exploring multi-modal frameworks that combine dermoscopic images, clinical photographs, and metadata represents a promising direction to better reflect real-world dermatological practice.

V. CONCLUSION

This review has surveyed CNN-based, hybrid, and transformer-based approaches for skin disease classification.

While CNNs remain effective on benchmark datasets, hybrid and transformer-based models offer promising alternatives for improved representation learning. Nevertheless, the challenges related to the dataset quality, reproducibility, and generalization persist. This highlights the need for transparent evaluation protocols and clinically reliable models.

REFERENCES

- [1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [3] H. L. Gururaj, N. Manju, A. Nagarjun, V. M. Aradhya, and F. Flammini, "Deepskin: a deep learning approach for skin cancer classification," *IEEE access*, vol. 11, pp. 50 205–50 214, 2023.
- [4] J. Mohan, A. Sivasubramanian, V. Ravi *et al.*, "Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable ai," *Computers in Biology and Medicine*, vol. 190, p. 110007, 2025.
- [5] international clinics, "International skin imaging collaboration (isic) archive," <https://www.isic-archive.com>, accessed: 2025.
- [6] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368*, 2019.
- [7] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.
- [8] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "Ph 2-a dermoscopic image database for research and benchmarking," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2013, pp. 5437–5440.
- [9] <https://www.nzdsi.org/>, "Dermnet nz: Dermatology resource and image library," <https://dermnetnz.org/>, accessed: 2025-XX-XX.
- [10] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: A review," *Neurocomputing*, vol. 464, pp. 364–393, 2021.
- [11] R. Saifan and F. Jubair, "Six skin diseases classification using deep convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, p. 3072, 2022.
- [12] K. Saranya, S. Vijayasharathi, N. Sasirekha, M. Rishika, and P. S. R. Rajeswari, "Skin disease detection using cnn (convolutional neural network)," in *2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)*. IEEE, 2024, pp. 1–6.
- [13] A. I. Kushartanto, R. T. Aldisa *et al.*, "Comparison of cnn and svm methods on web-based skin disease classification process," *Sinkron: jurnal dan penelitian teknik informatika*, vol. 8, no. 2, pp. 778–788, 2024.
- [14] A. Harbola, N. Danu, D. Negi, A. Gupta, A. Joshi, and R. Negi, "Unveiling dermatology: A deep learning approach to skin disease detection with convolutional neural network," in *Challenges in Information, Communication and Computing Technology*. CRC Press, 2025, pp. 459–464.
- [15] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets—a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, 2022.
- [16] M. K. Hasan, M. T. E. Elahi, M. A. Alam, M. T. Jawad, and R. Martí, "Dermoexpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation," *Informatics in Medicine Unlocked*, vol. 28, p. 100819, 2022.
- [17] A. Kumar and K. K. Tiwari, "Disease classification in dermatology: A cnn-rf hybrid approach," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIoT)*. IEEE, 2024, pp. 1–5.
- [18] A. Ansari, A. Singh, M. Singh, and V. Kukreja, "Enhancing skin disease classification: A hybrid cnn-svm model approach," in *2024 International Conference on Automation and Computation (AUTOCOM)*. IEEE, 2024, pp. 29–32.

TABLE III: summaray table of vision transformers approches

Reference	Year	Dataset	Model	segmentation	augmentation	Metrics	Result
[29]	2022	HAM10000	Vit	-	Yes	Acc	$\approx 94\%$
[25]	2023	HAM10000	Vit	-	Oversampling	Acc	94.1%
[28]	2023	Multiple datasets	Hybrid Transformer + SVM	-	Yes	Acc	$\approx 91\%$
[29]	2023	ISIS 2019 + Private	ConvNext + SwinTransformer	-	Yes	Acc	$\approx 91\%$
[27]	2023	ISIC2020 + PH2+ Others	BEiT (Transformer)	-	-	F1-Score	92.0%
[30]	2023	ISIC 2018 + Private data	Vit	-	Yes	Acc	93.81%
[31]	2024	HAM10000 + Clinical data	Vit	-	Yes	Acc	99.3%
[22]	2024	ISIC 2019	Vit + SwinVit	YOLOv8	Yes	Acc Precision	99.97% 100%
[24]	2024	ISIC 2020	MBVit	-	GAN	AUC	0.986%
[23]	2025	DermNet (+ healthy skin)	Swin Transformer	-	Yes	Acc	97.20%
[4]	2025	DermNet + ISIC 2018	Swin Transformer Vit DinoV2	-	Yes	Acc	93.26% 92.35% 97.27%

- [19] V. Kumar, D. Banerjee, R. Chauhan, S. Kukreti, and K. S. Gill, "Optimizing skin disease classification: A comprehensive investigation of cnn-svm hybrid models with layer configurations," in *2024 3rd International Conference for Innovation in Technology (INOCON)*. IEEE, 2024, pp. 1–6.
- [20] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [21] —, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [22] D. K. Saha, A. M. Joy, and A. Majumder, "Yotransvit: A transformer and cnn method for predicting and classifying skin diseases using segmentation techniques," *Informatics in Medicine Unlocked*, vol. 47, p. 101495, 2024.
- [23] J. Ahmad, M. U. Farooq, F. Zafeer, N. Shahid *et al.*, "Classification of 24 skin conditions using swin transformer: leveraging dermnet & healthy skin dataset," in *International Conference on Energy, Power, Environment, Control and Computing (ICEPECC 2025)*, vol. 2025. IET, 2025, pp. 511–519.
- [24] R. Wang, X. Chen, X. Wang, H. Wang, C. Qian, L. Yao, and K. Zhang, "A novel approach for melanoma detection utilizing gan synthesis and vision transformer," *Computers in Biology and Medicine*, vol. 176, p. 108572, 2024.
- [25] G. Yang, S. Luo, and P. Greer, "A novel vision transformer model for skin cancer classification," *Neural Processing Letters*, vol. 55, no. 7, pp. 9335–9351, 2023.
- [26] C. Xin, Z. Liu, K. Zhao, L. Miao, Y. Ma, X. Zhu, Q. Zhou, S. Wang, L. Li, F. Yang *et al.*, "An improved transformer network for skin cancer classification," *Computers in Biology and Medicine*, vol. 149, p. 105939, 2022.
- [27] S. Vachmanus, T. Noraset, W. Piyanonpong, T. Rattananukrom, and S. Tuarob, "Deepmetaforge: A deep vision-transformer metadata-fusion network for automatic skin lesion classification," *IEEE Access*, vol. 11, pp. 145 467–145 484, 2023.
- [28] S. Khan and A. Khan, "Skinvit: A transformer based method for melanoma and nonmelanoma classification," *Plos one*, vol. 18, no. 12, p. e0295151, 2023.
- [29] S. Hao, L. Zhang, Y. Jiang, J. Wang, Z. Ji, L. Zhao, and I. Ganchev, "Convnext-st-aff: A novel skin disease classification model based on fusion of convnext and swin transformer," *IEEE Access*, vol. 11, pp. 117 460–117 473, 2023.
- [30] G. Cai, Y. Zhu, Y. Wu, X. Jiang, J. Ye, and D. Yang, "A multimodal transformer to fuse images and metadata for skin disease classification," *The Visual Computer*, vol. 39, no. 7, pp. 2781–2793, 2023.
- [31] S. Remya, T. Anjali, and V. Sugumaran, "A novel transfer learning framework for multimodal skin lesion analysis," *IEEE Access*, 2024.