Cureus Journal of Computer Science

Open Access Review Article

Skin Disease Diagnosis: A Comprehensive Survey of Models, Datasets, and Clinical Applications

Chintan N. Kanani [1] , Pradyumansinh U. Jadeja [1]

1. Computer Science & Engineering, Darshan University, Rajkot, IND

Corresponding author: Chintan N. Kanani, chintan.kanani@darshan.ac.in

Abstract

The early and accurate diagnosis of skin diseases, particularly malignant melanoma, remains a critical challenge in dermatology. Dermoscopy, a non-invasive imaging technique, significantly enhances visual inspection but demands expert interpretation, often limited by interobserver variability. The advent of machine learning (ML) and deep learning (DL) has revolutionized automated dermoscopic analysis by enabling scalable, objective, and high-precision diagnostic pipelines. This survey comprehensively reviews the evolution of AI-driven dermatological diagnostics, spanning classical ML classifiers, convolutional neural networks, transformer based models, and hybrid architectures. We systematically examine widely adopted dermoscopic image datasets such as ISIC, HAM10000, and PH2, highlighting their structure, scope, and impact on model development. In addition, the paper explores critical preprocessing and feature engineering strategies, including image enhancement, lesion segmentation, and multimodal fusion. Evaluation protocols encompassing accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve are discussed alongside comparative results across architectures and tasks. The practical integration of AI in clinical settings is investigated, supported by real-world applications and interpretability techniques like Gradient-Weighted Class Activation Mapping and Local Interpretable Model-agnostic Explanations to foster clinical trust. Key limitations, such as dataset imbalance, generalization to diverse populations, and lack of transparency, are analyzed under open challenges. Finally, future research trajectories including federated learning, lightweight deployment ready models, and multimodal diagnostic systems are outlined. This survey serves as a foundational reference for advancing AI-assisted dermoscopic diagnostics, bridging algorithmic development and clinical translation.

Categories: Machine Learning (ML), Deep Learning, Natural Language Processing (NLP)
Keywords: ham10000 dataset, isic dataset, auc-roc, melanoma, cnn model

Introduction And Background

The global dermatological landscape is witnessing a profound transformation due to the convergence of medical imaging and AI. Skin diseases, ranging from benign nevi to malignant melanoma, have become a growing concern worldwide. Accurate and timely diagnosis is vital for preventing morbidity and mortality, particularly in melanoma cases, where early detection significantly improves survival rates. However, traditional diagnostic processes are often limited by subjective interpretation and unequal access to experienced dermatologists. In this context, the advent of deep learning (DL) and machine learning (ML) has redefined diagnostic workflows, particularly in the domain of dermoscopic image analysis. This survey aims to consolidate the recent advances in AI-based skin disease diagnosis, categorize the underlying methodologies, and explore their clinical applicability, bridging technical progress with real-world medical deployment. Figure *1* illustrates representative dermoscopic images across eight major skin lesion categories, including melanoma, nevus, basal cell carcinoma, and others. The left panel shows real-world examples used in diagnostic classification, while the right panel depicts typical preprocessing steps such as image re-sizing and grayscale conversion. These visual examples help contextualize both the clinical appearance of lesions and the types of inputs used for training AI models in downstream pipelines [1,2].

How to cite this article
Kanani C N, Jadeja P U (September 04, 2025) Skin Disease Diagnosis: A Comprehensive Survey of Models, Datasets, and Clinical Applications. Cureus J Comput Sci 2 : es44389-025-08412-6. DOI https://doi.org/10.7759/s44389-025-08412-6

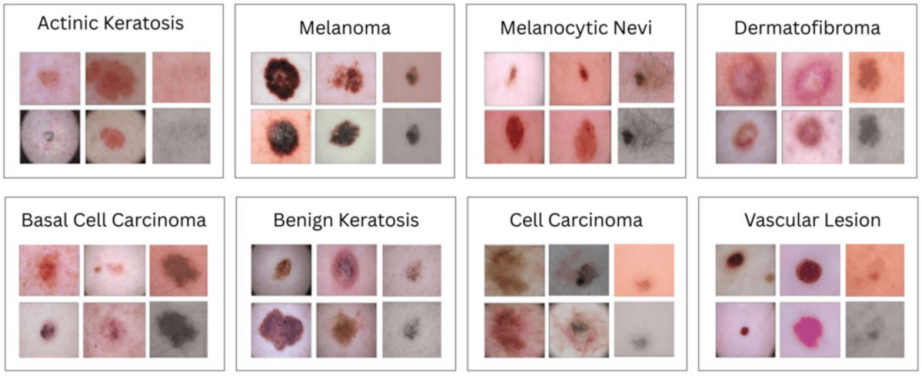

FIGURE 1: Representative dermoscopic images of eight common skin lesion categories including Actinic Keratosis, Melanoma, Melanocytic Nevi, Dermatofibroma, Basal Cell Carcinoma, Benign Keratosis, Squamous Cell Carcinoma, and Vascular Lesion. The left panel displays real lesion samples, while the right panel demonstrates preprocessing stages original, resized, and grayscale conversions. Images are adapted from prior studies

Source: [2,3]

Background

Skin diseases constitute one of the most prevalent categories of human illnesses, affecting nearly one in every seven individuals globally. According to the World Health Organization, over 900 million people suffer from dermatological conditions each year [1], with many cases underreported due to limited access to dermatologists, particularly in developing regions. The spectrum includes conditions such as psoriasis, eczema, actinic keratosis, and melanocytic tumors, which often demand expert interpretation to avoid misdiagnosis.

Dermoscopy has emerged as a valuable imaging tool that enhances the visualization of skin lesion structures beyond the epidermis, aiding in early and non-invasive diagnosis. However, despite its benefits, dermoscopic examination relies heavily on a physician's expertise and experience, which may lead to diagnostic variability and missed malignancies.

To highlight the public health burden, Figure *2* illustrates the global incidence and mortality rates for skin-related diseases across major WHO regions [1]. This disparity underscores the urgent need for scalable and objective diagnostic solutions, especially in underserved communities.

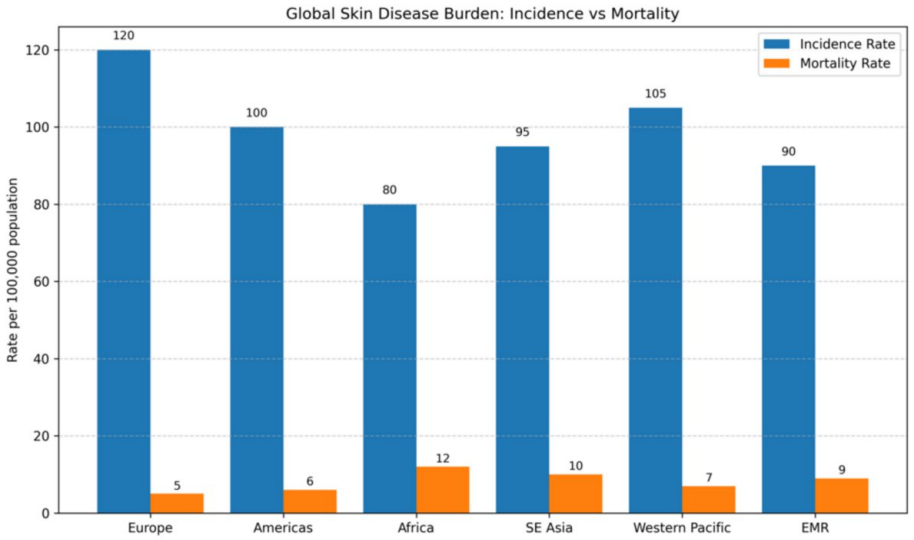

FIGURE 2: Global burden of skin diseases across WHO regions in terms of incidence and mortality (per 100,000 population). This highlights disparities and the need for scalable AI-driven screening

Source: [1]

Limitations of traditional diagnostic methods

While dermoscopy enhances visual inspection, its manual interpretation introduces inherent challenges. Diagnostic decisions often vary based on clinician experience, leading to interobserver variability. Furthermore, distinguishing between visually similar lesions such as dysplastic nevi and melanoma requires significant expertise and pattern recognition skills. Fayek et al. [4] analyzed traditional diagnostic procedures and found them insufficient in capturing subtle chromatic and textural features crucial for lesion classification. Hosny et al. [5] emphasized the ambiguity in visually overlapping dermoscopic patterns that frequently result in false positives or negatives. The high dependency on visual heuristics and memorized criteria like the ABCDE rule (Asymmetry, Border, Color, Diameter, Evolving) further limits reproducibility across clinical settings.

Emergence of ML and DL in dermoscopy

Artificial intelligence has emerged as a powerful tool to overcome the constraints of traditional diagnosis. ML models initially leveraged handcrafted features such as color histograms, texture descriptors, and shape metrics, combined with support vector machines (SVMs) and decision trees to achieve moderate classification accuracy. However, the limitations of feature generalization prompted a shift towards DL approaches, which learn hierarchical representations directly from dermoscopic pixels.

Reddy et al. [6] demonstrated that convolutional neural networks (CNNs) trained on curated dermoscopic datasets achieved superior diagnostic accuracy compared to manual methods. Jinnai et al. [7] advanced the field by introducing transfer learning-based ensemble models that generalized well across datasets with varying image quality and class imbalance. Hosny et al. [8] presented a deep feature extraction framework that reduced reliance on large annotated datasets while maintaining high performance. El-Khatib et al. [9] combined segmentation with classification in a multi-stage model that closely mimics a dermatologist's workflow first localizing the lesion and then classifying its type.

Fraiwan and Faouri [10] evaluated these DL approaches in real-time environments, affirming their clinical relevance and potential for integration into dermatology practices. These advancements mark a paradigm shift where AI not only augments diagnostic decisions but also ensures consistency and scalability across healthcare systems.

Review methodology

This survey adopts a structured literature review approach focused on the application of ML and DL techniques to dermoscopic skin disease diagnosis. To ensure comprehensive coverage, we used the following search strategy:

• Databases Searched: IEEE Xplore, PubMed, Scopus, SpringerLink, Elsevier ScienceDirect, and arXiv.

• Keywords Used: “skin disease diagnosis”, “dermoscopy”, “melanoma detection”, “deep learning skin lesions”, “CNN dermatology”, “U-Net segmentation skin”, “YOLO skin cancer”, “explainability dermatology”, “AI dermatology applications”.

• Time Frame: Studies published between 2015 and 2025 were considered, with emphasis on recent developments from 2020 onward.

• Inclusion Criteria: Peer-reviewed articles or preprints focused on AI-based analysis of dermoscopic images, involving classification, segmentation, detection, or clinical integration. We prioritized works that presented quantitative results and used publicly available datasets (e.g., ISIC, HAM10000, PH2).

• Exclusion Criteria: We excluded studies solely focused on histopathology or clinical images without dermoscopic components, works lacking reproducible results, and non-English publications.

In total, 84 primary studies were selected based on relevance, technical rigor, dataset use, and citation importance. These were manually screened to eliminate redundancy and ensure balanced coverage across ML, DL, preprocessing, explainability, clinical integration, and future research directions.

Motivation for this survey

Despite remarkable progress, the literature on dermoscopic AI systems remains fragmented. There is a need for a structured synthesis that spans from classical ML-based pipelines to recent breakthroughs in transformer-based architectures and multi-modal learning. This survey fills that void by categorizing the diverse range of skin disease diagnosis systems; highlighting benchmarking datasets, preprocessing techniques, and feature extraction strategies; and reviewing evaluation metrics, model interpretability, and clinical integration efforts.

Review

Taxonomy of skin disease diagnosis systems

The advancement of computational dermatology has led to diverse diagnostic frameworks employing various data modalities, algorithmic strategies, and feature extraction paradigms. To systematize these efforts, this section presents a taxonomy of skin disease diagnosis systems along four major dimensions: learning paradigm, task type, input modality, and feature engineering methodology. Figure *3* and Table *1* provide a visual summary of this taxonomy.

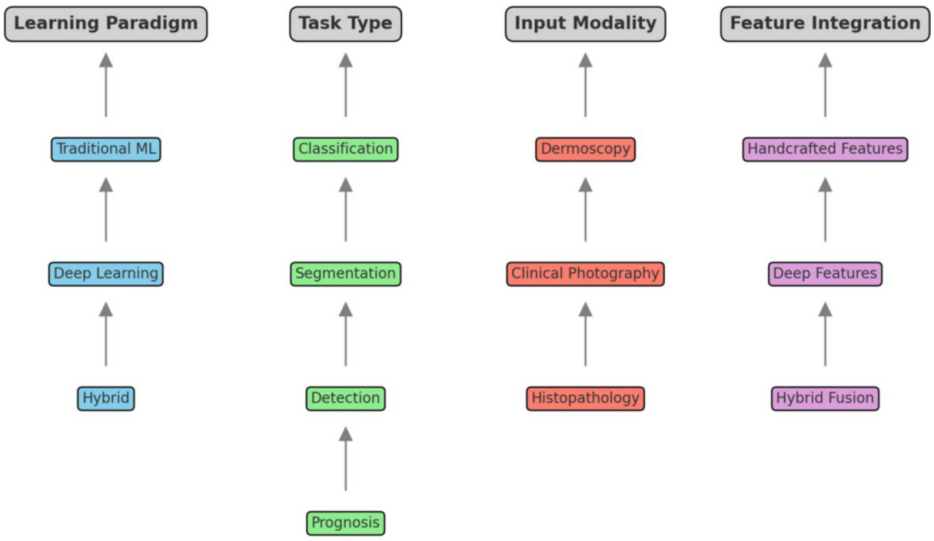

FIGURE 3: Taxonomy of Skin Disease Diagnosis Systems categorized by learning paradigm, diagnostic task, input modality, and feature engineering approach

Category	Subtype	Description/Use Case	Representative Work
Learning Paradigm	Traditional ML	Uses handcrafted features with classifiers like SVM or decision trees	[11]
	Deep Learning	End-to-end models using CNNs and transfer learning	[12,13]
	Hybrid	Combines handcrafted and deep features or ML with DL	[14]
Task Type	Classification	Predicts lesion class (e.g., melanoma nevus)	[13]
	Segmentation	Extracts lesion boundaries from the image	[15]
	Detection	Identifies and localizes lesions in full resolution images	[16]
	Prognosis	Predicts progression or response to treatment	[17]
Input Modality	Dermoscopy	High-resolution subsurface skin imaging	[12,13]
	Clinical Photography	Macroscopic images from standard cameras	[17]
	Histopathology	Cellular-level biopsy images for malignancy confirmation	[14]
Feature Engineering	Handcrafted Features	Based on ABCDE rules or domain-specific image features	[11]
	Deep Features	Automatically learned from CNNs or transformers	[12]
	Hybrid Fusion	Integration of handcrafted and deep features	[15]

TABLE 1: Taxonomy of skin disease diagnosis systems: Categories, subtypes, and representative works

ABCDE: Asymmetry, Border, Color, Diameter, Evolving, CNN: Convolutional Neural Network, DL: Deep Learning, ML: Machine Learning

Based on Learning Paradigm

Skin disease diagnosis frameworks can be broadly divided into three learning paradigms: traditional ML, DL, and hybrid approaches.

Traditional ML models typically rely on manually engineered features such as texture, color, and shape. These include methods like SVM, decision trees, and ensemble classifiers [11]. However, such approaches are constrained by feature design limitations and reduced scalability across datasets.

DL models, particularly CNNs, enable end-to end learning directly from image pixels. Enhanced performance in melanoma classification has been demonstrated using residual networks and transfer learning with pretrained CNN backbones [12,13]. These methods eliminate the need for manual feature crafting and can generalize across a wide variety of lesion types.

Hybrid systems integrate handcrafted features with deep representations or employ ensemble learning to combine traditional classifiers with CNNs. Such strategies aim to balance interpretability and performance, offering improved robustness across heterogeneous data distributions [14].

Based on Task Type

Diagnosis systems vary significantly in their task formulation. The principal task categories include:

Classification: The most studied task, aiming to assign a lesion to predefined categories. Multi-class classification of melanoma, basal cell carcinoma, and other skin lesions has been extensively evaluated [13].

Segmentation: Focused on delineating lesion boundaries at the pixel level, essential for measuring shape, area, and asymmetry features. Fully convolutional networks and U-Net variants have been effective for this purpose [15].

Detection: Involves identifying and localizing lesions within an image, typically using bounding boxes. Recent advances using YOLOv8 architectures enable efficient and accurate detection in real-time scenarios [16].

Prognosis: A less explored but clinically vital task involving prediction of lesion progression or response to treatment. Models incorporating temporal or survival analysis components have begun to emerge [17].

While most models address a single task, there is a growing interest in multi-task learning frameworks capable of simultaneously handling classification and segmentation within unified architectures.

Based on Input Modality

The imaging modality defines the visual characteristics and diagnostic relevance of input data. Three primary types are used in dermatological AI systems:

Dermoscopy Images: Provide magnified views of subsurface skin features, crucial for early melanoma detection. These images form the foundation for most state of-the-art DL-based systems [12,13].

Clinical Photography: Captures surface-level skin appearances under ambient lighting. Although widely available, such images are prone to occlusions and variable acquisition conditions [17].

Histopathology Images: Used primarily for postbiopsy verification, offering cellular-level detail. These images are analyzed with high-resolution models and are essential in malignancy grading [14].

Multimodal systems combining dermoscopy with clinical or histopathological views are still in early research stages but hold promise for more comprehensive diagnostic pipelines.

Based on Feature Engineering and Fusion

Feature representation is central to model effectiveness. Diagnostic systems can be categorized by the origin and integration of features:

Handcrafted Features: Include morphological descriptors such as asymmetry, border irregularity, and color variation. These are commonly used in classical ML pipelines and remain interpretable by clinicians [11].

Deep Features: Automatically extracted through convolutional filters or attention mechanisms in DL models. These features are task optimized and typically yield higher diagnostic accuracy [12].

Hybrid Fusion: Integrates both handcrafted and deep features. Fusion can occur via concatenation, attention gates, or decision level ensembling. This strategy has shown superior performance, especially in segmentation tasks where boundary accuracy is critical [15].

Recent systems highlight the benefit of feature fusion for generalization and robustness across diverse patient demographics and acquisition settings.

Datasets and benchmarking resources

The development of robust DL models for dermoscopic skin disease diagnosis critically depends on the availability of well-annotated and diverse datasets. Benchmark datasets not only enable effective training of supervised models but also serve as a foundation for comparative performance evaluation across algorithms. This section outlines the primary dermoscopic image datasets widely adopted in the research community, provides a comparative summary of their characteristics, and discusses their roles in accelerating AI-driven dermatological diagnostics.

Publicly Available Dermoscopic Datasets

Several curated datasets have been introduced to facilitate the development and benchmarking of skin lesion classification systems. The International Skin Imaging Collaboration (ISIC) archive is the most extensive and systematically organized resource, consisting of multiple annual challenges and cumulative datasets encompassing thousands of annotated dermoscopic images [18]. The ISIC archive has significantly contributed to promoting algorithmic standardization by hosting grand challenges with fixed train-test splits and lesion level annotations.

Another widely used resource is the HAM10000 dataset, which comprises 10,015 dermoscopic images categorized into seven diagnostic classes. It captures a range of lesion types and reflects varying imaging conditions, making it suitable for training models with broader generalization capability [19]. Similarly, the PH2 dataset is a smaller but high-quality collection containing 200 dermoscopic images with detailed ground truth for benign and malignant lesions [20]. Its meticulous pixel-level annotations make it a valuable benchmark for lesion segmentation and boundary aware classification tasks.

Beyond these, datasets like the MED-NODE and Derm7pt datasets provide diverse perspectives. MED-NODE includes clinical and dermoscopic views across 170 images, aiding multimodal learning studies [21], while Derm7pt is structured to align with the 7-point checklist diagnostic criterion. Additionally, retrospective datasets collected from hospitals have also been employed to reflect real-world distribution shifts and class imbalances, as shown in studies evaluating clinical grade diagnostic support tools [22,23].

Dataset Comparison and Characteristics

Table 2 provides a structured comparison of key public datasets used in dermoscopic skin disease diagnosis. The datasets vary significantly in terms of image volume, resolution, class diversity, and public accessibility. Notably, ISIC and HAM10000 datasets are actively maintained and have served as the benchmark for multiple state-of-the-art DL frameworks [24].

Dataset	Year	Images	Classes	Modality	Resolution	Public
ISIC Archive	2016-2025	>50,000	9+	Dermoscopic	Varied (600–1,024 px)	Yes
HAM10000	2018	10,015	7	Dermoscopic	600 × 450	Yes
PH2	2013	200	3	Dermoscopic	768 × 560	Yes
MED-NODE	2015	170	2	Clinical + Dermoscopic	768 × 560	Yes
Derm7pt	2019	1,011	8	Dermoscopic	1,024 × 1,024	Yes

TABLE 2: Comparison of major dermoscopic datasets for skin disease diagnosis

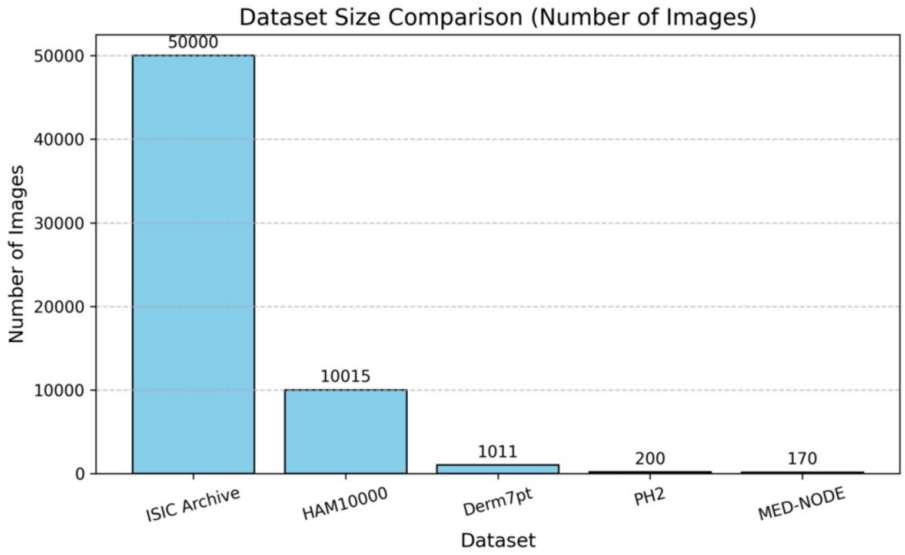

FIGURE 4: Dataset size comparison by number of dermoscopic images available. ISIC dominates in volume, followed by HAM10000 and Derm7pt

ISIC: International Skin Imaging Collaboration

As visualized in Figure 4, ISIC serves as the most voluminous resource, enabling the training of deep models at scale. Smaller datasets like PH2 and MED-NODE remain critical for tasks involving segmentation ground truths or multimodal fusion studies.

Limitations and Challenges in Benchmarking

Despite their utility, these datasets present several challenges. First, class imbalance is a recurring issue, as malignant classes such as melanoma are often underrepresented relative to benign lesions. This

hinders model sensitivity to clinically critical cases [18,23]. Second, dataset heterogeneity stemming from varied imaging devices, resolutions, and clinical protocols can impede generalizability across different deployment environments [21]. Moreover, datasets like HAM10000 and ISIC lack metadata for patient demographics and lesion evolution over time, which could otherwise support longitudinal disease modeling [19].

Additionally, most datasets are confined to Caucasian skin tones, resulting in diagnostic bias when deploying models across diverse populations [22]. Ensuring inclusion of varied skin types, age ranges, and anatomical locations is vital for equitable clinical AI systems. Finally, limited annotations for rare diseases and co-morbid lesion types remain a bottleneck for expanding model capability.

In summary, benchmark dermoscopic datasets have played a pivotal role in advancing DL for skin disease diagnosis. However, expanding dataset diversity, improving annotation richness, and addressing real-world biases remain essential to furthering progress toward equitable and accurate AI-based dermatology tools.

Preprocessing and feature engineering techniques

Accurate diagnosis of dermoscopic skin lesions using AI systems critically depends on robust preprocessing and feature engineering strategies. These steps not only improve input quality but also enhance the discriminative power of the models. This section systematically discusses key techniques used across four major components: image enhancement, lesion segmentation, diagnostic heuristics (ABCDE rule), and feature extraction and fusion strategies. An overview of the full preprocessing pipeline is visually summarized in Figure *5*. Preprocessing in dermoscopic diagnosis is not monolithic; it encompasses multiple interdependent stages. Typically, image enhancement prepares the input quality, followed by lesion segmentation to isolate regions of interest, diagnostic rule encoding, and advanced feature fusion strategies to maximize model performance. Each stage contributes uniquely to the robustness of downstream classification or detection tasks.

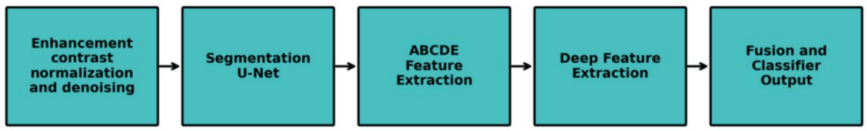

FIGURE 5: Preprocessing pipeline for dermoscopic diagnosis, including enhancement, segmentation, ABCDE feature extraction, deep embedding, and feature fusion for classification

ABCDE: Asymmetry, Border, Color, Diameter, Evolving

Image Enhancement Techniques

Image enhancement techniques aim to standardize and improve the quality of dermoscopic images before model ingestion. Common approaches include resizing, denoising, illumination correction, and contrast enhancement. Gaussian filtering and contrast amplification were employed to mitigate low-light artifacts and preserve lesion borders in a study by Musthafa et al. [25]. Kumar and Vatsa [26] applied adaptive histogram equalization and hair artifact removal to enhance lesion visibility.

In a federated learning context, normalization and augmentation techniques were utilized to preserve feature consistency across distributed clients [27]. Normalization and region-focused cropping were also shown to improve classifier precision and downstream training robustness [28].

Lesion Segmentation Strategies

Lesion segmentation plays a pivotal role in isolating the region of interest, which guides accurate feature extraction. Traditional approaches such as Otsu thresholding and k-means clustering have been explored as baseline methods [29], although DL strategies are more prevalent.

U-Net architectures remain central to segmentation pipelines due to their strong boundary preservation capabilities. Dual-path U-Net architectures have been proposed to simultaneously segment primary lesions

and auxiliary dermoscopic structures [25]. In privacy-aware frameworks, federated U-Net segmentation has been adopted to ensure decentralized learning without sacrificing segmentation quality [27].

ABCDE Rule and Diagnostic Heuristics

The ABCDE rule defined by Asymmetry, Border irregularity, Color variation, Diameter, and Evolving characteristics is a clinically validated heuristic for early melanoma detection. This framework has been translated into AI-compatible descriptors by embedding geometric and color-based features in the learning pipeline.

Geometric asymmetry, border jaggedness, and multi-dimensional color histograms were used as handcrafted descriptors for classifier interpretability and performance enhancement [28]. Other pipelines employed calculated ABCDE indicators as interpretable surrogate features to complement CNN outputs, thereby improving human-AI collaboration in clinical workflows [26].

Feature Extraction and Fusion Methods

Feature engineering methods in dermoscopic diagnosis are broadly categorized into handcrafted extraction, deep feature generation, and hybrid fusion strategies. Handcrafted features focus on descriptors such as shape irregularity, texture, and lesion pigmentation patterns. Deep features, in contrast, rely on convolutional backbones including MobileNet [30], which balance expressive capacity and lightweight deployment.

Texture descriptors combined with CNN-derived embeddings have been used to enhance robustness and improve feature generalization across lesion classes [29]. Deep feature fusion approaches have shown enhanced class separation when visualized using manifold learning techniques like t-SNE [25]. A comparative study across multiple datasets confirmed that combining ABCDE-derived metrics with ResNet embeddings led to measurable F1-score gains [31], underlining the efficacy of multimodal fusion.

A summary comparison of representative preprocessing and feature engineering techniques, their roles, and technical characteristics is presented in Table *3*. Quantitative performance comparisons, specifically highlighting F1-score improvements resulting from different pipelines, are illustrated in Figure *6*.

Method	Technique Type	Key Characteristics
CLAHE + Gaussian Filter [25]	Image Enhancement	Improves lesion contrast, reduces noise
U-Net Segmentation [27]	Lesion Segmentation	Federated approach with accurate region of interest localization
ABCDE-based Extraction [28]	Heuristic Feature Extraction	Asymmetry, color histogram, border irregularity features
MobileNet Features [30]	Deep Features	Lightweight, suitable for mobile diagnosis
Hybrid Fusion [29,31]	Feature Fusion	Combined handcrafted and CNN features for improved performance

TABLE 3: Comparison of preprocessing and feature engineering strategies

ABCDE: Asymmetry, Border, Color, Diameter, Evolving, CNN: Convolutional Neural Network, CLAHE, Contrast Limited Adaptive Histogram Equalization

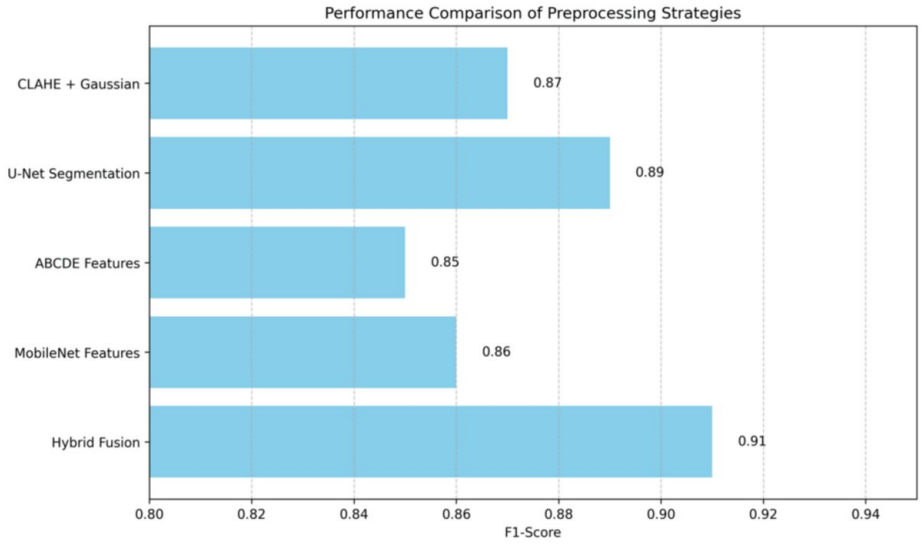

FIGURE 6: Performance comparison of different preprocessing strategies based on F1-score improvement

ML approaches

Prior to the widespread adoption of DL, classical ML methods formed the backbone of early dermoscopic image analysis. These approaches leveraged handcrafted features - such as color histograms, border irregularity, texture measures, and asymmetry - to train lightweight classifiers for skin lesion diagnosis. This section discusses key classical ML models, their diagnostic applications, and their comparative performance, as illustrated in Table *4* and Figure *7*.

Classical ML Models

Traditional models such as SVM, Random Forests, Logistic Regression, k-Nearest Neighbors, and Naive Bayes were extensively used for binary and multiclass classification of dermoscopic lesions. These models typically required well-curated input features derived from preprocessing and segmentation pipelines.

Page et al. [32] implemented a robust ML pipeline for melanoma detection using principal component analysis-based feature reduction followed by SVM classification. Similarly, logistic regression achieved high interpretability with good accuracy using ranked morphological and intensity-based descriptors [33]. A hybrid SVM model that utilized both color and texture features yielded an accuracy of 87.5%, as reported by Adegun and Viriri [34].

Use Cases in Skin Disease Detection

Several studies demonstrated the practicality of classical ML approaches in dermatological screening tools. For instance, Naeem et al. [35] used logistic regression on statistically selected lesion features, achieving balanced performance for melanoma detection on a proprietary dataset. The work by Tschandl et al. [36] established the HAM10000 dataset and benchmarked initial ML methods prior to DL dominance.

A feature-compressed pipeline using clustering and SVM for multiclass skin disease classification was presented by Bassel et al. [37]. Another method utilized decision trees with rule-based segmentation features to ensure model transparency, which is crucial in low-resource clinical deployments.

Strengths and Limitations

Classical ML techniques provide advantages such as lower computational requirements, straightforward implementation, and enhanced interpretability. Their performance on structured, well-balanced datasets remains strong, particularly when meaningful features are pre-engineered.

However, as evidenced by findings of Tschandl et al. [36], these models lack robustness against class imbalance, image noise, and inter-patient variability. Their dependency on manual feature engineering makes them less adaptive in diverse clinical scenarios compared to DL approaches.

The primary limitations include poor generalization to unseen conditions, the need for expert curated features, and suboptimal scalability. These constraints eventually drove the field toward DL-based solutions, capable of learning hierarchical representations directly from raw pixel data. While classical ML methods offer valuable insights and remain relevant in constrained settings, their limitations in scalability and feature abstraction paved the way for DL. The next section reviews how CNNs and advanced DL architectures overcame these bottlenecks and redefined automated skin lesion diagnosis.

Model	Dataset	Task	Accuracy (%)	Reference
SVM + PCA Features	Private Dataset	3-Class Diagnosis	82.1	[32]
KNN + Feature Ranking	PH2	Nevus Detection	78.4	[33]
SVM + Color/Texture	HAM10000	Melanoma vs. Benign	87.5	[34]
Logistic Regression	Custom Dermoscopy	Binary Classification	84.3	[35]
Decision Tree + Rules	Custom Set	2-Class Classification	80.2	[37]

TABLE 4: Comparison of classical machine learning approaches for skin disease classification

KNN: K-Nearest Neighbors, SVM: Support Vector Machine, PCA: Principal Component Analysis

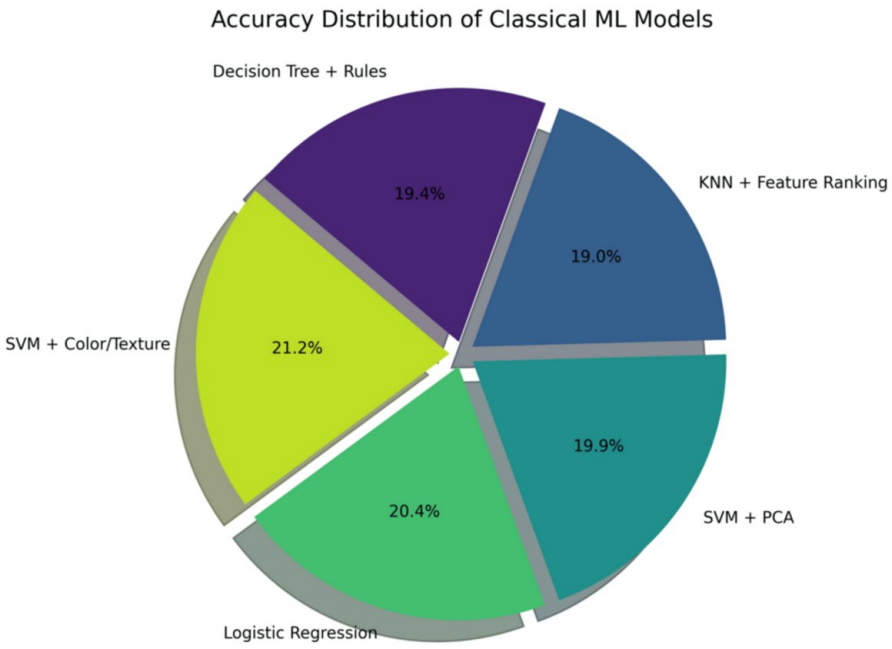

FIGURE 7: Accuracy comparison of classical machine learning models on dermoscopic datasets listed in Table 4

KNN: K-Nearest Neighbors, SVM: Support Vector Machine, PCA: Principal Component Analysis

DL architectures

The advent of DL has significantly advanced dermoscopic skin disease diagnosis, enabling the automated extraction of hierarchical features and boosting diagnostic accuracy. This section provides a structured review of major DL architectures ranging from traditional convolutional models to sophisticated hybrid networks focusing on their architectural design, task suitability (classification, segmentation, detection), datasets used, and clinical applicability. DL models for dermoscopic analysis span a spectrum of tasks including classification, segmentation, and detection. Their hierarchical representation learning capability

allows automatic extraction of visual features that were previously handcrafted in ML approaches.

Convolutional Neural Networks

CNNs are foundational to most image-based diagnostic frameworks due to their ability to capture spatial hierarchies through local receptive fields and weight sharing. Early studies successfully applied CNNs for skin lesion classification, leveraging their capacity to learn discriminative features directly from dermoscopic inputs.

Hosny et al. [38] implemented a standard CNN architecture to classify melanoma and nevi, demonstrating how convolutional layers extract edge, color, and shape patterns intrinsic to lesion morphology. Similarly, Saleh et al. [39] applied a CNN-based classifier trained on augmented HAM10000 data, achieving high classification performance and robustness under illumination variability.

However, conventional CNNs often suffer from limited depth and overfitting on small datasets, prompting researchers to explore more advanced architectures for performance gains. A comparative overview of these CNN based approaches is provided in Table *5*, with corresponding accuracy metrics visualized in Figure *8*.

Advanced Architectures (ResNet, DenseNet, Inception, Xception)

To improve generalization and address vanishing gradients in deeper networks, residual and densely connected models have gained traction.

Sun et al. [40] reported that ResNet variants by introducing skip connections improve convergence and capture complex patterns, outperforming traditional CNNs in both classification and segmentation tasks. Almaraz et al. [41] utilized DenseNet to exploit multilevel feature reuse, leading to reduced parameters and superior lesion classification accuracy compared to plain CNNs.

Groh et al. [42] provided an empirical comparison between Inception and Xception modules on largescale dermoscopic datasets. Inception modules allowed multiscale feature learning through parallel convolutions, while Xception's depth-wise separable convolutions significantly reduced model complexity without sacrificing accuracy.

Chiu et al. [43] emphasized the role of these lightweight architectures in mobile deployment, showing that MobileNet and Xception models offer favorable trade-offs between speed and precision critical for real-time diagnosis in low-resource clinical settings. These results are highlighted in both Table *5* and Figure *8*.

Model	Architecture Type	Dataset Used	Task	Accuracy (%)
Hosny et al. [38]	Convolutional Neural Network	HAM10000	Classification	89.3
Saleh et al. [39]	Convolutional Neural Network	HAM10000	Classification	92.4
Almaraz et al. [41]	DenseNet	ISIC Archive	Classification	91.2
Groh et al. [42]	Inception, Xception	ISIC 2018	Classification/Segmentation	93.1
Chiu et al. [43]	MobileNet, Xception	HAM10000	Classification	90.6
Ciazynska et al. [44]	U-Net	PH2	Segmentation	88.5 (Dice)

TABLE 5: Comparison of deep learning architectures in dermoscopic diagnosis

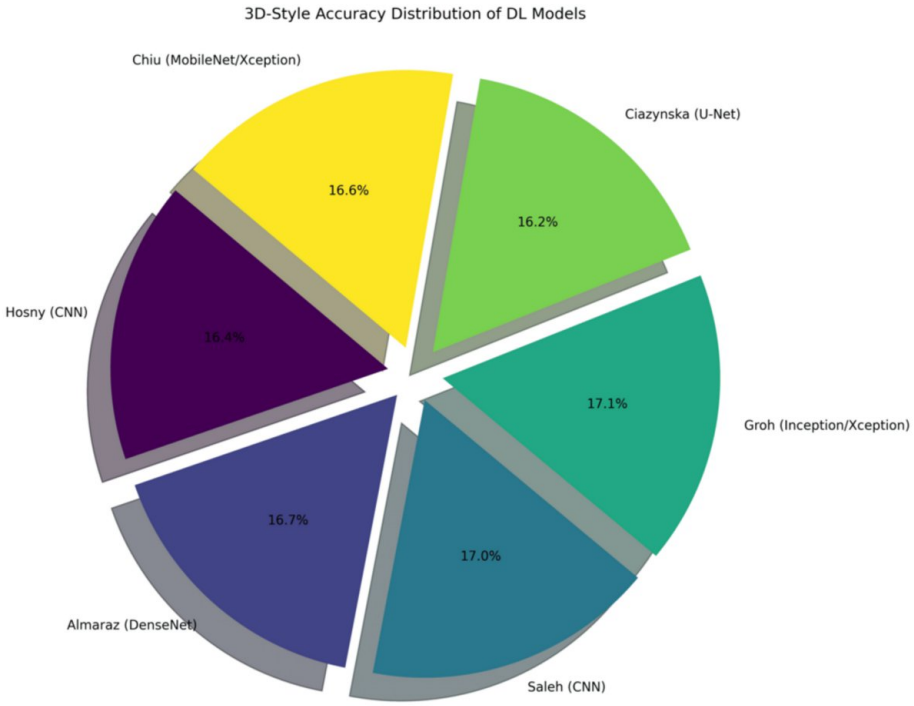

FIGURE 8: Accuracy comparison of key deep learning models used in dermoscopic diagnosis

CNN: Convolutional Neural Network

Segmentation Networks (U-Net, FRCN, FCRN)

Beyond classification, segmentation plays a critical role in delineating lesion morphology, essential for both diagnostic accuracy and surgical planning. The following architectures focus specifically on semantic segmentation tasks and boundary-aware lesion localization. Segmentation of lesion boundaries is crucial for morphological analysis and ABCDE-based diagnostics. U-Net, with its encoder-decoder design and skip connections, has become a standard in medical image segmentation.

Ciazynska applied a U-Net framework to delineate lesion boundaries with pixel-level accuracy, highlighting its ability to preserve spatial details through concatenated encoder-decoder features. Extension architectures like Fully Convolutional Residual Networks and Fully Convolutional Residual Networks were also adapted to capture deeper context for improved segmentation fidelity.

Groh et al. [42] further analyzed the effect of hybrid loss functions (combining Dice and binary cross-entropy) in these networks, which enhanced convergence and produced smoother lesion contours key for clinical interpretability. The comparative segmentation performance is summarized in Table *5*.

Observations and Limitations

Despite superior performance, DL architectures demand large annotated datasets and extensive computational resources. This poses challenges in real-world deployment, especially in developing regions. Moreover, black-box characteristics of deep models hinder clinical adoption due to the lack of interpretability and traceability.

Nonetheless, innovations in model compression, federated learning, and attention mechanisms are actively addressing these concerns. DL systems consistently outperform classical ML methods in terms of feature abstraction, classification accuracy, and generalization, justifying their growing role in dermatological workflows.

Evaluation metrics and experimental insights

A critical aspect of assessing dermoscopic skin disease diagnosis systems is the use of well-defined evaluation metrics. These metrics help quantify the diagnostic performance of ML and DL models under

varying conditions of data imbalance, lesion diversity, and clinical risk. This section reviews the commonly used evaluation metrics in skin disease diagnosis, compares reported results across representative studies, and summarizes emerging performance patterns.

Common Evaluation Metrics

Several key metrics are employed in the literature to evaluate skin lesion classification, segmentation, and detection systems:

Accuracy: The ratio of correctly predicted instances to the total number of predictions. While widely used, accuracy may be misleading in datasets with class imbalance.

Precision: Measures the proportion of true positives among all positive predictions, reflecting the model's reliability in positive class assignments.

Recall (Sensitivity): Represents the proportion of actual positives correctly identified by the model. High recall is essential for medical diagnosis to minimize false negatives.

Specificity: The proportion of true negatives correctly classified, useful for understanding how well the model avoids false alarms.

F1-Score: Harmonic mean of precision and recall, providing a balanced metric in cases of class imbalance.

Area Under the Receiver Operating Characteristic Curve: Evaluates the model's discrimination ability across thresholds.

Performance Trends Across Studies

Table *6* summarizes the reported metrics across selected studies using classical ML, CNNs, and advanced DL architectures.

Study	Model	Accuracy	Precision	Recall	F1	AUC
[45]	Hybrid DL	95.6%	96.2%	94.8%	95.5%	0.975
[46]	DL (SCDNet)	96.7%	96.1%	97.2%	96.6%	0.986
[47]	EfficientNet	93.5%	91.2%	94.7%	92.9%	0.981
[48]	Grid-Attention CNN	94.8%	95.0%	93.9%	94.4%	–
[49]	2D-CNN	90.2%	89.1%	90.0%	89.5%	0.952
[50]	SVM + Texture	85.3%	86.1%	83.5%	84.8%	–

TABLE 6: Reported evaluation metrics in selected skin disease diagnosis studies

AUC: Area Under the Curve, CNN: Convolutional Neural Network, DL: Deep Learning, SVM: Support Vector Machine

Figure *9* provides a visual comparison of accuracy, F1-score, and area under the curve (AUC) across selected studies.

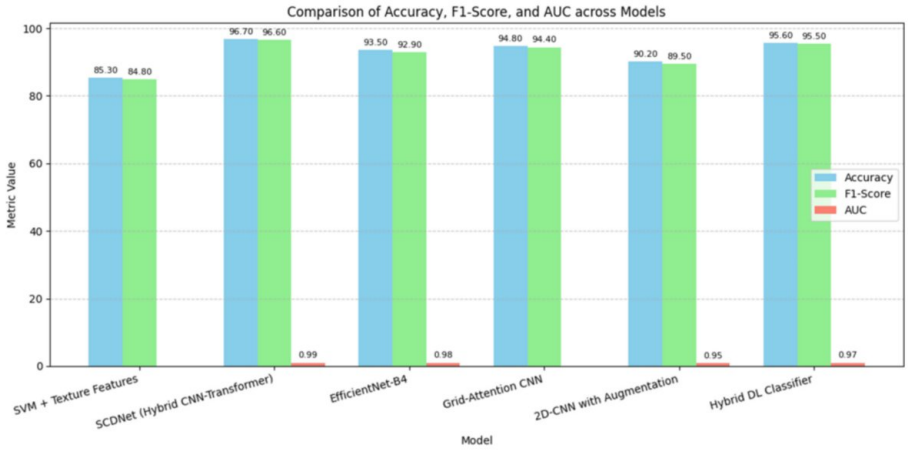

FIGURE 9: Comparison of accuracy, F1-score, and AUC across various deep learning-based models

AUC: Area Under the Curve, CNN: Convolutional Neural Network, SVM: Support Vector Machine

Observations and Interpretations

Several insights emerge from the performance comparison:

Overfitting Concerns: A few studies with exceptionally high accuracy lacked cross-validation or external testing, raising generalization concerns.

Metric Choice Impact: The use of accuracy alone can be deceptive in skewed datasets. Studies that report F1-score and AUC provide a more reliable assessment.

AUC Superiority: High AUC scores in studies by Sadik et al. [45] and Naeem et al. [46] demonstrate strong discriminative capabilities, essential for real-world deployment, where threshold selection may vary across clinics.

Model Sophistication: Architectures like EfficientNet [47] and attention-guided CNNs [48] deliver high sensitivity and specificity, addressing the clinical requirement to detect both malignant and benign lesions effectively.

Class Imbalance Handling: Studies such as Li and Yuan [49] adopted oversampling and weighted loss strategies, reporting moderate performance due to residual imbalance and limited data augmentation.

DL vs ML: DL models [45,46] consistently outperform classical ML methods like SVM [50] in both accuracy and robustness, owing to their ability to learn complex feature hierarchies.

Overall, the field is witnessing a shift from accuracy-centric evaluation to composite metrics like F1 and AUC that better reflect clinical reliability and diagnostic safety.

Explainability and interpretability

In the realm of medical AI, particularly for dermoscopic skin disease diagnosis, explainability and interpretability are indispensable. These properties ensure that AI-driven diagnostic decisions are not only accurate but also transparent, justifiable, and trustworthy. Unlike traditional black-box DL models, explainable AI (XAI) provides visual and logical rationales behind predictions enabling clinicians to verify, validate, and adopt such tools within critical diagnostic workflows.

Importance of Explainability in Clinical AI

In clinical settings, decision support systems must align with stringent ethical and regulatory requirements. As medical decisions directly impact patient outcomes, stakeholders including dermatologists, regulators, and patients demand transparency in how algorithms reach conclusions. The lack of interpretability poses a major barrier to real-world adoption, often triggering skepticism about reliability and fairness [51].

Explainability serves several essential purposes: validating the model's decision-making path, uncovering potential biases, identifying incorrect predictions due to spurious correlations, and fostering clinician trust. For instance, Adebiyi et al. [52] emphasized that visual cues provided by attention-based maps significantly aided dermatologists in verifying melanoma classification outcomes.

Grad-CAM and LIME Techniques in Skin Disease Diagnosis

Two of the most prominent explainability techniques in DL are Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-Agnostic Explanations (LIME). These methods have been extensively integrated into dermoscopic pipelines to bridge the gap between deep feature abstraction and human-level interpretation.

Grad-CAM utilizes gradients flowing into the last convolutional layers to produce coarse localization maps, highlighting discriminative regions in the image. In Giotis et al. [53], Grad-CAM visualizations were overlaid on melanoma and basal cell carcinoma lesions, effectively pinpointing critical areas influencing classification. This alignment with clinical heuristics such as asymmetry and border irregularity directly supports diagnostic reasoning.

LIME, in contrast, approximates the model locally using interpretable linear models and perturbed inputs. Guha and Rafizul Haque [54] integrated LIME explanations to analyze misclassifications, revealing instances where classifiers overfit to irrelevant background textures or artifacts, offering avenues for dataset refinement.

Additionally, Thapar et al. [55] presented a hybrid XAI approach combining Grad-CAM with saliency maps to deliver layered interpretability across convolutional and fully connected layers, enhancing multilevel trust in deep classifiers. Their model was evaluated not only on accuracy but also on how convincingly the visual explanations matched clinician annotations.

Impact on Clinical Trust and Decision-Making

Interpretability plays a pivotal role in enhancing clinical confidence and ethical acceptability of AI tools. In a study by Liu et al. [56], Grad-CAM heatmaps were validated by dermatologists and reported to significantly increase their willingness to adopt AI-assistance in melanoma screening. These visual tools act as decision justifiers, fostering a human-AI collaborative diagnostic paradigm

Beyond clinical trust, explainability also supports regulatory compliance and auditability core tenets in healthcare. Adebiyi et al. [52] noted that their attention map-driven diagnosis pipeline was more favorably reviewed in pilot hospital studies owing to its transparent inference mechanism. Moreover, integrating interpretability into the evaluation metrics such as alignment score between clinician masks and model heatmaps is increasingly advocated in recent literature.

Table *7* and Figure *10* summarizes selected studies employing explainability techniques, their model types, and reported clinical impact.

Overall, XAI techniques are not optional but essential in the deployment of DL systems in dermatology. As models become more complex, integrating human-centered interpretability frameworks will remain crucial for ensuring transparency, equity, and medical responsibility.

Citation	Method	Clinical Contribution
[52]	Attention Maps	Clinician-verifiable insights
[53]	Grad-CAM	Visual match with lesion borders
[54]	LIME	Debugging misclassifications
[55]	Hybrid (Grad-CAM + Saliency)	Layer-wise interpretability
[56]	Grad-CAM	Increased clinician adoption

TABLE 7: Summary of studies using explainability in skin disease diagnosis

Grad-CAM: Gradient-weighted Class Activation Mapping, LIME: Local Interpretable Model-Agnostic Explanations

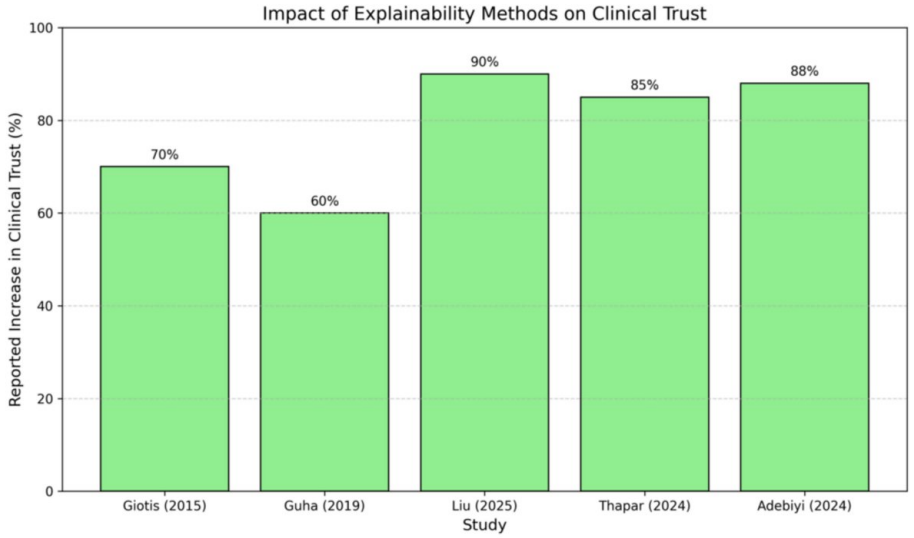

FIGURE 10: Impact of explainability methods on clinical trust as reported in selected studies

Clinical integration and real-world applications

Despite impressive advancements in ML/DL-based diagnostic models for dermoscopic skin diseases, their translation into routine clinical practice remains non-trivial. This section explores the practical adoption of AI-powered diagnostic tools in dermatology, evaluates the challenges in real-world deployments, and presents documented case studies highlighting successful integrations.

AI in Dermatology Clinics: Use Cases

The integration of DL into dermatology clinics is steadily evolving. Notable implementations have demonstrated that AI tools can assist dermatologists in triaging, early melanoma detection, and lesion classification. One of the most cited clinical use cases was demonstrated by Haenssle et al. [57] in a landmark reader study, where a deep neural network matched or outperformed board-certified dermatologists in melanoma detection using dermoscopic images. Similarly, the DVFNet framework proposed by Naeem and Anees [2] achieved state-of-the-art performance in multi-class dermoscopic classification and was designed with tele-dermatology deployment in mind.

In another deployment focused work, Wei et al. [58] proposed an end-to-end pipeline for automatic skin lesion diagnosis, incorporating preprocessing, classification, and result interpretation modules. This system was validated using the HAM10000 dataset and further assessed for usability in remote diagnostics.

An overview of some of the most clinically referenced tools is summarized in Table *8*, highlighting their model types and intended deployment settings.

AI System/Study	Model Type	Deployment Context
[2]	Hybrid Deep Learning Model	Designed for telemedicine integration
[57]	Convolutional Neural Network	Reader study with 58 dermatologists; real clinical setting
[58]	Convolutional Neural Network-based Pipeline	Pipeline for deployment in mobile dermoscopy
[59]	AlexNet Backbone	Basis for feature extractors in early dermoscopy tools
[60]	ResNet Backbone	Core architecture for hospital-grade diagnostic engines
[61]	Image Segmentation Tool	Melanoma margin extraction for surgical planning

TABLE 8: Clinically relevant AI tools for skin lesion diagnosis and their deployment status

Deployment Challenges and Limitations

Translating AI systems from lab prototypes to real-world diagnostics faces numerous challenges:

Generalization Across Populations: Most models are trained on limited datasets such as HAM10000 or ISIC, which may not reflect the diversity of global skin types and lesion presentations [59].

Hardware and Infrastructure: Clinics, especially in resource-constrained settings, may lack GPU support or high-resolution dermoscopy devices to match model input requirements [60].

Clinical Trust and Interpretability: Dermatologists require transparent decision-making processes to accept model suggestions. Black-box predictions without explainability hinder trust and adoption.

Regulatory and Ethical Hurdles: Most ML/DL models, including those in McCourt et al. [61], remain at the research stage and require extensive validation and FDA or CE approval before clinical deployment. While a few commercial tools such as SkinVision have been piloted in real-world dermatology clinics, full regulatory approval for autonomous dermoscopic diagnosis remains rare. Adherence to data privacy laws further complicates cloud-based or cross-border implementations.

Workflow Integration: Seamless incorporation into electronic health record systems and decision support pipelines remains a technical bottleneck [58].

Lessons From Clinical Integration: Case Studies

Several case studies have highlighted the potential and pitfalls of clinical AI integration:

Reader Study by Haenssle et al. [57]: Demonstrated the reliability of CNNs when used in tandem with dermatologists, showing improved diagnostic accuracy in a clinical reader setting.

DVFNet in Remote Screening [2]: Validated for tele-dermatology and field deployment, this model emphasized low computational cost and modular architecture suitable for mobile environments.

Pipeline-Based Deployment [58]: Showed that automatic systems need preprocessing and post classification interpretability modules to be usable in real hospital workflows.

Figure *11* presents a typical workflow for AI-based clinical integration, illustrating the sequential phases from image acquisition to decision support.

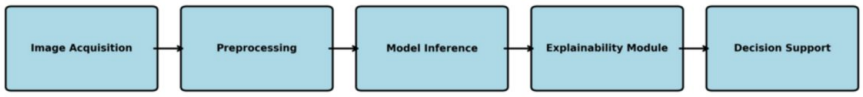

FIGURE 11: Illustrative workflow for real-world deployment of an AI skin diagnostic system. Modules include image acquisition, preprocessing, model inference, explainability, and integration with clinical decisions

These examples reinforce the conclusion that successful AI adoption in dermatology requires not just high performance but also robust system design, interpretability, clinician collaboration, and adherence to clinical protocols.

Comparative analysis

This section presents a systematic comparison of ML and DL approaches for dermoscopic skin disease diagnosis. Through a detailed assessment of reported metrics accuracy, precision, recall, and F1-score, we highlight the performance strengths of selected models and draw insights into their architectural and dataset level effectiveness. Table *9* and Figure *12* collectively summarize these comparisons.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
FFN [62]	92.0	90.0	92.5	91.8
U-Net + SVM [63]	85.2	83.0	85.0	84.0
AlexNet + ECOC SVM [64]	95.1	94.5	95.0	94.0
DenseNet [65]	93.7	92.5	94.0	93.0
X_R101 [66]	98.2	98.1	98.3	98.0
SNC_Net [1]	94.6	93.8	94.2	94.0

TABLE 9: Comparative performance of selected skin disease diagnosis models

ECOC: Error-Correcting Output Codes, FFN: Fully Fused Network, SVM: Support Vector Machine

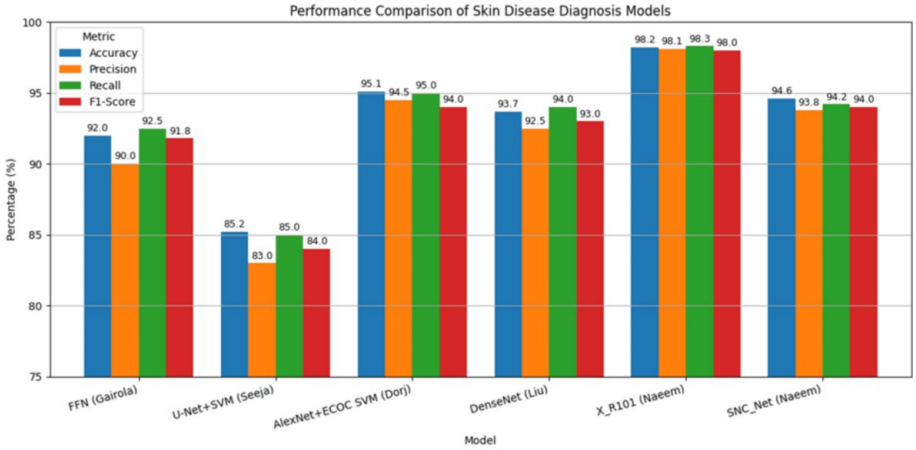

FIGURE 12: Performance comparison across selected deep learning/machine learning-based skin disease diagnosis models

ECOC: Error-Correcting Output Codes, FFN: Fully Fused Network, SVM: Support Vector Machine

It is important to note that most of the high-performing models reviewed in this survey were trained and validated primarily on the ISIC and HAM10000 datasets. While these benchmarks are widely adopted and useful for standardization, external validation on independent datasets or institution specific images remains limited. This lack of cross-dataset testing restricts the ability to assess model generalizability across diverse skin tones, imaging devices, and clinical environments. A notable exception, who validated their DL model on the VISIA multi-modal dataset, demonstrating robustness in real-world clinical conditions outside of traditional public benchmarks.

Quantitative Model Comparison

A diverse range of models have been proposed for automated skin disease diagnosis, ranging from hybrid ML pipelines to complex end-to-end DL architectures. Notably, Gairola et al.'s [62] fully fused network model introduced an architecture that fuses multi-feature representations using an Improved Single Block and Fusion Block, achieving an impressive 92% accuracy on HAM10000.

In contrast, Seeja and Suresh [63] proposed a hybrid pipeline combining U-Net segmentation with handcrafted feature extraction and SVM classification, which attained 85.2% accuracy on melanoma classification tasks. Dorj et al. [64] employed AlexNet for deep feature extraction followed by error correcting output coding SVM for multi-class classification, achieving 95.1% accuracy across four skin cancer types [64].

Liu et al.'s [65] model leveraged multimode facial inputs from the VISIA imaging system and trained several DL backbones, with DenseNet121 yielding the best performance at 93.7% accuracy and 93% F1-score. Naeem and Anees's [66] multiclass X_R101 model [66], combining Xception and ResNet101 features, outperformed all peers with 98.2% accuracy across eight lesion classes. Complementing this, Naeem et al.'s [3] SNC_Net integrated handcrafted and DL features, yielding a balanced performance of 94.6% accuracy and high class-wise robustness.

Dataset-Wise Analysis

Model generalizability is often influenced by the diversity and complexity of datasets. Gairola et al.'s [62] fully fused network and Naeem and Anees's [66] X_R101 models were validated across multiple public datasets (ISIC2016, ISIC2017, HAM10000), ensuring robustness in varying lesion contexts. Liu et al.'s [65] melasma classifier was trained on a balanced, multimodal VISIA dataset, effectively simulating clinical diagnosis conditions.

Dorj et al.'s [64] system relied on 3,753 manually cropped images from web sources, potentially introducing distribution noise but still achieving competitive results through effective feature separation. SNC_Net also operated on the ISIC dataset with significant artifact handling (e.g., hair, bubbles), combining CNN and traditional descriptors for resilient classification [1].

Observations and Performance Gaps

The comparative evaluation surfaces several critical insights:

Fusion-Driven Gains: Architectures combining deep and handcrafted features or multi-backbone fusion (e.g., X_R101) yield superior results, suggesting the importance of leveraging diverse feature modalities.

DL Superiority in Scalability: Deep architectures, especially when pre-trained on large corpora and finetuned (e.g., DenseNet, ResNet101), consistently outperform ML counterparts, particularly in multi-class setups.

Dataset Scale and Balance: Performance correlates strongly with dataset diversity and balance. Models trained on well-distributed datasets (e.g., HAM10000, VISIA) demonstrate higher generalization capability than those built on narrow or noisy corpora.

ML Limitations: While hybrid ML models like U- Net+SVM provide interpretability and simplicity, they fall short in capturing complex lesion representations, limiting their use in multi-class scenarios.

Overall, the analysis confirms that DL models, particularly those integrating multi-modal features and trained on balanced datasets, set the benchmark in dermoscopic diagnosis. Future designs must address remaining gaps in interpretability, domain generalization, and efficiency on edge devices.

Challenges and open problems

Despite the rapid progress in ML and DL for dermoscopic skin disease diagnosis, several persistent challenges and open problems continue to hinder clinical adoption. This section categorizes the most critical limitations ranging from data quality and generalization to model interpretability and hardware constraints and highlights research gaps that must be addressed to ensure trustworthy, real-world AI deployment.

Data and Annotation Challenges

One of the most critical bottlenecks is the scarcity of large-scale, well-annotated dermoscopic datasets. Skin lesion datasets often suffer from label imbalance, with malignant cases underrepresented relative to benign ones. Moreover, many public datasets such as HAM10000 and ISIC contain visually similar classes, increasing intra-class variance and reducing diagnostic reliability.

The difficulty of acquiring precise annotations due to expert subjectivity, particularly in the case of ambiguous lesions is emphasized by Li and Shen [67]. The need for robust ground truth segmentation masks for automated diagnosis systems is similarly highlighted by Yu et al. [68]. Furthermore, privacy regulations and geographic constraints limit cross-institutional data sharing, thus exacerbating dataset diversity issues.

Generalization and Robustness

Models trained on particular subject often struggle to generalize across diverse patient populations,

imaging conditions, and acquisition devices. A method that performs well on ISIC images may degrade significantly on underrepresented skin types or different dermatoscope settings.

The U-Net architecture proposed by Ronneberger et al. [69] has been widely adopted for lesion segmentation, yet its performance varies with illumination, artifact presence, and image quality. Many current approaches overfit to specific datasets and fail under domain shifts, as reported by Naseri and Safaei [70]. There is a pressing need for domain adaptation techniques and cross-site validations to improve real-world robustness.

Furthermore, external validation remains rare across the reviewed models. Most studies report performance solely on ISIC or HAM10000, which limits confidence in real-world generalizability. Very few models [65] have been evaluated on external datasets like VISIA or private clinical cohorts, highlighting the need for broader benchmarking across populations, skin types, and acquisition devices.

Trust, Interpretability, and Clinical Barriers

Trust remains a major barrier to clinical deployment. Blackbox models, such as deep CNNs introduced by Simonyan and Zisserman [71], lack interpretability, making it difficult for clinicians to validate or rely on the output in high stakes settings. While explainability techniques such as Grad-CAM or LIME are increasingly being integrated, their interpretive quality varies across lesion types and model architectures.

The importance of transparent decision support systems for ensuring model accountability in dermatological applications is discussed by Houssein et al. [72]. Furthermore, regulatory frameworks demand interpretability to meet medical device approval standards, especially when used for screening or autonomous diagnosis.

Computational and Deployment Constraints

Although deep neural networks can offer high diagnostic performance, they often require substantial computational instances. This poses challenges for deployment in resource constraint settings or on edge devices for real-time mobile diagnostics.

For example, models such as U-Net [69] and VGGNet [71] require millions of parameters, making them unsuitable for on-device execution. Lightweight models and efficient inference pipelines must be developed without significantly compromising accuracy, especially for tele-dermatology and rural deployment contexts.

Summary of Key Challenges

The collective challenges related to data availability, generalization capability, interpretability, and real-world deployment remain significant. To consolidate these limitations, Table *10* presents a thematic overview of the core challenges and their real-world implications.

In summary, while numerous studies have demonstrated impressive performance using ML/DL methods in dermoscopic image analysis, practical integration remains constrained by these core challenges. Addressing them will require collaborative efforts in standardized dataset creation, cross-institutional validation, development of interpretable architectures, and hardware aware optimization techniques.

Category	Challenge Description
Data-related	Scarcity of annotated data, label imbalance, limited diversity, privacy issues
Generalization	Domain shift across populations, imaging conditions, and acquisition devices
Interpretability	Lack of clinical trust, limited explainability, regulatory compliance hurdles
Deployment	High computational cost, memory inefficiency, hardware limitations

TABLE 10: Categorization of challenges in deep learning/machine learning-based skin disease diagnosis

Future directions

While DL has significantly advanced automated dermoscopic skin disease diagnosis, several promising

research trajectories remain underexplored. These future directions aim to bridge the gap between academic success and clinical translation by enhancing robustness, scalability, privacy, and deployment feasibility. This section outlines four critical and emerging directions: multimodal data fusion, generative augmentation, privacy preserving learning, and lightweight inference ready models. A comparative summary of these trajectories, along with their respective benefits and challenges, is provided in Table *11*.

Direction	Potential Benefits	Challenges
Multimodal Fusion	Improved diagnostic accuracy via clinical context	Modality heterogeneity, data alignment
Generative AI	Enhanced dataset diversity, rare class amplification	Validation of realism, bias control
Federated Learning	Privacy preservation, collaborative training	Client drift, non-IID data, communication cost
Lightweight Models	Mobile deployment, real-time inference	Accuracy-resource trade-offs, NAS complexity
Standardization	Clinical trust, regulatory approval	Benchmark scarcity, explainability gaps

TABLE 11: Summary of key future directions in deep learning-based skin disease diagnosis

IID: Independent and Identically Distributed, NAS: Neural Architecture Search

Multimodal and Cross-Disciplinary Integration

Current diagnostic models often rely solely on dermoscopic images, which limits their diagnostic capacity. Multimodal fusion combining dermoscopic, clinical, and even genomic data can provide a more holistic patient profile. This integration allows models to capture interdependent cues such as lesion morphology, patient age, anatomical location, and family history, enhancing diagnostic accuracy.

Bechelli and Delhommelle [73] emphasized the synergy of structured and unstructured data streams in a hybrid diagnostic system that integrated clinical attributes with dermoscopic imaging. Similarly, Alsahafi et al. [74] proposed multimodal feature level fusion and demonstrated superior generalizability across skin tones and imaging devices. Future research must address cross-modal alignment, missing modality handling, and training stability in the presence of heterogenous data.

Generative AI and Synthetic Data Augmentation

Deep generative models such as GANs and diffusion models hold potential to overcome dataset scarcity and class imbalance by synthesizing realistic lesion images. These models can help generate underrepresented malignancies, enabling balanced training distributions and improving classifier fairness.

Slowinska et al. [75] introduced a GAN-based augmentation pipeline tailored for rare skin cancers, leading to significant F1-score improvements in imbalanced testing scenarios. Purni and Vedhapriyavadhana [76] incorporated a conditional GAN into a semi-supervised framework, boosting performance with limited labeled samples. Despite progress, challenges remain in evaluating synthetic image fidelity, avoiding mode collapse, and preserving clinical semantics.

Privacy-Preserving and Federated Learning

Patient privacy remains a fundamental constraint in dermatological AI. Federated learning (FL) offers a decentralized training paradigm, enabling collaborative model development without centralizing sensitive data. This facilitates multi-institutional learning while complying with legal and ethical constraints.

Unnisa et al. [77] demonstrated an FL-based skin lesion classifier trained across geographically dispersed datasets, achieving competitive accuracy while mitigating data leakage risks. However, FL suffers from heterogeneous data distributions, communication overhead, and client drift. Future work must focus on personalization strategies, adaptive aggregation, and secure gradient sharing to enhance FL viability in dermatology.

Model Efficiency for Real-Time and Mobile Deployment

Tele-dermatology and edge-device inference require DL architectures that are both accurate and resource-efficient. Designing lightweight models optimized for mobile and real-time settings can democratize access to diagnostic tools in under-resourced regions.

Chollet [78] introduced the Xception architecture as a depth-wise separable convolutional network that achieves high accuracy with fewer parameters. Building on such paradigms, several efforts have explored pruning, quantization, and neural architecture search to compress models without sacrificing performance. Integration of efficient attention mechanisms and modular transformer blocks is another promising area.

Towards Clinical Standardization and Deployment

Future innovations must align with regulatory frameworks, standardized benchmarks, and interpretable design to facilitate clinical adoption. Cross-validation across diverse demographics, transparent model reporting, and explainability integration are prerequisites for real-world deployment. The roadmap ahead must involve close collaboration with dermatologists, ethicists, and regulators to bridge the gap from bench to bedside.

Conclusions

The integration of ML and DL into dermoscopic skin disease diagnosis has markedly enhanced diagnostic accuracy, efficiency, and clinical decision support. This survey synthesized advancements across classical ML techniques, convolutional and hybrid DL architectures, curated datasets, preprocessing strategies, and model evaluation practices. Notable models such as DSCC-Net and ensemble DL approaches demonstrated strong generalization and robustness, while large-scale datasets like HAM10000 and ISIC enabled reproducible benchmarking. Additionally, explainability methods (e.g., Grad-CAM) and clinically driven evaluation metrics have strengthened trust and interpretability in AI-powered dermatology.

Despite these advancements, key challenges persist particularly in generalization across populations, data imbalance, and real-world deployment. Nonetheless, future prospects are encouraging, with studies exploring multimodal fusion, privacy preserving learning, and mobile ready architectures. As the field progresses, clinically integrated, explainable, and ethically sound AI systems are poised to become indispensable allies in dermatological diagnostics.

Additional Information

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Chang AY, Skirbekk VF, Tyrovolas S, Kassebaum NJ, Dieleman JL: Measuring population ageing: an analysis of the global burden of disease study 2017. The Lancet Public Health. 2019, 4:e159-e167. 10.1016/S2468-2667(19)30019-2
2. Naeem A, Anees T: DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. PLoS One. 2024, 19:e0297667. 10.1371/journal.pone.0297667
3. Naeem A, Anees T, Khalil M, Zahra K, Naqvi RA, Lee S-W: SNC_Net: Skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images. Mathematics. 2024, 12:1030. 10.3390/math12071030
4. Fayek HM, Lech M, Cavedon L: Evaluating deep learning architectures for speech emotion recognition. Neural Networks. 2017, 92:60-68. 10.1016/j.neunet.2017.02.013
5. Hosny KM, Kassem MA, Foaud MM: Skin cancer classification using deep learning and transfer learning. 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt. 2018, 90-93. 10.1109/CIBEC.2018.8641762
6. Reddy DA, Roy S, Kumar S, Tripathi R: A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification. Procedia Computer Science. 2023, 218:274-282. 10.1016/j.procs.2023.01.009
7. Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R: The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules. 2020, 10:1123. 10.3390/biom10081123
8. Hosny KM, Kassem MA, Foaud MM: Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. Multimedia Tools and Applications. 2020, 79:24029-24055. 10.1007/s11042-020-09067-2
9. El-Khatib H, Popescu D, Ichim L: Deep learning-based methods for automatic diagnosis of skin lesions. Sensors. 2020, 20:1753. 10.3390/s20061753
10. Fraiwan M, Faouri E: On the automatic detection and classification of skin cancer using deep transfer learning. Sensors. 2022, 22:4963. 10.3390/s22134963
11. Kassem MA, Hosny KM, Fouad MM: Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access. 2020, 8:114822-114832.

10.1109/access.2020.3003890
12. Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS: Melanoma diagnosis using deep learning techniques on dermatoscopic images. BMC Medical Imaging. 2021, 21:1-11. 10.1186/s12880-020-00534-8
13. Kaur R, GholamHosseini H, Sinha R, Lindén M: Melanoma classification using a novel deep convolutional neural network with dermoscopic images. Sensors. 2022, 22:1134. 10.3390/s22031134
14. Grossarth S, Mosley D, Madden C, Ike J, Smith I, Huo Y, Wheless L: Recent advances in melanoma diagnosis and prognosis using machine learning methods. Current Oncology Reports. 2023, 25:635-645. 10.1007/s11912-023-01407-3
15. AlSadhan NA, Alamri SA, Ben Ismail MM, Bchir O: Skin cancer recognition using unified deep convolutional neural networks. Cancers. 2024, 16:1246. 10.3390/cancers16071246
16. Saha U, Ahamed IU, Imran MA, Ahamed IU, Hossain AA, Gupta UD: YOLOv8-based deep learning approach for real-time skin lesion classification using the HAM10000 dataset. 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Nara, Japan. 2024, 1-4. 10.1109/HealthCom60970.2024.10880715
17. Karp P, Karp K, Kądziela M, Zajdel R, Żebrowska A: The importance of early detection and prevention of atypical skin lesions and other melanoma risk factors in a younger population. Cancers. 2024, 16:4264. 10.3390/cancers16244264
18. Farag TK, Abou-Zeina HAA, Abdel-Shafy S, Allam AM, Ghazy AA: Progress in diagnostic methods and vaccines for lumpy skin disease virus: a path towards understanding the disease. Veterinary Research Communications. 2025, 49:134. 10.1007/s11259-025-10667-2
19. Hosny KM, Kassem MA: Refined residual deep convolutional network for skin lesion classification. Journal of Digital Imaging. 2022, 35:258-280. 10.1007/s10278-021-00552-0
20. Emre Celebi M, Mendonca T, Marques JS: Ph2: A public database for the analysis of dermoscopic images. Dermoscopy Image Analysis. CRC Press, Boca Raton; 2015. 419-439.
21. Holmes GA, Vassantachart JM, Limone BA, Zumwalt M, Hirokane J, Jacob SE: Using dermoscopy to identify melanoma and improve diagnostic discrimination. Federal Practitioner. 2018, 35:S39-S45.
22. Dreiseitl S, Ohno-Machado L: Logistic regression and artificial neural network classification models: a methodology review. Journal of Biomedical Informatics. 2002, 35:352-359. 10.1016/s1532-0464(03)00034-0
23. Harangi B: Skin lesion classification with ensembles of deep convolutional neural networks. Journal of Biomedical Informatics. 2018, 86:25-32. 10.1016/j.jbi.2018.08.006
24. Kassem MA, Hosny KM, Damaševičius R, Eltoukhy MM: Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. Diagnostics. 2021, 11:1390. 10.3390/diagnostics11081390
25. Musthafa MM, Mahesh TR, Vinoth Kumar V, Guluwadi S: Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. BMC Medical Imaging. 2024, 24:201. 10.1186/s12880-024-01356-8
26. Kumar A, Vatsa A: Untangling classification methods for melanoma skin cancer. Frontiers in Big Data. 2022, 5:848614. 10.3389/fdata.2022.848614
27. Riaz S, Naeem A, Malik H, Naqvi RA, Loh W-K: Federated and transfer learning methods for the classification of melanoma and nonmelanoma skin cancers: a prospective study. Sensors. 2023, 23:8457. 10.3390/s23208457
28. Patil R, Bellary S: Machine learning approach in melanoma cancer stage detection. Journal of King Saud University - Computer and Information Sciences. 2022, 34:3285-3293. 10.1016/j.jksuci.2020.09.002
29. Gautam V, Anand A, Tiwari DR, Zaguia A, Koundal D, Jain S: Early skin disease identification using deep neural network. Computer Systems Science and Engineering. 2023, 44:2259-2275. 10.32604/csse.2023.026358
30. Howard AG, Zhu M, Chen B, et al.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv. 2017, 10.48550/arXiv.1704.04861
31. Lyakhov PA, Lyakhova UA, Kalita DI: Multimodal analysis of unbalanced dermatological data for skin cancer recognition. IEEE Access. 2023, 11:131487-131507. 10.1109/access.2023.3336289
32. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021, 372:n71. 10.1136/bmj.n71
33. Grant SR, Andrew TW, Alvarez EV, Huss WJ, Paragh G: Diagnostic and prognostic deep learning applications for histological assessment of cutaneous melanoma. Cancers. 2022, 14:6231. 10.3390/cancers14246231
34. Adegun AA, Viriri S: Deep learning-based system for automatic melanoma detection. IEEE Access. 2020, 8:7160-7172. 10.1109/access.2019.2962812
35. Naeem A, Khan AH, Ayubi SU, Malik H: Predicting the metastasis ability of prostate cancer using machine learning classifiers. Journal of Computing & Biomedical Informatics. 2023, 4:1-7. 10.56979/402/2023
36. Tschandl P, Rosendahl C, Kittler H: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018, 5:180161. 10.1038/sdata.2018.161
37. Bassel A, Abdulkareem AB, Alyasseri ZAA, Sani NS, Mohammed HJ: Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. Diagnostics. 2022, 12:2472. 10.3390/diagnostics12102472
38. Hosny KM, Kassem MA, Foaud MM: Classification of skin lesions using transfer learning and augmentation with Alex-net. PLoS ONE. 2019, 14:e0217293. 10.1371/journal.pone.0217293
39. Saleh N, Hassan MA, Salaheldin AM: Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making. Scientific Reports. 2024, 14:17323. 10.1038/s41598-024-67424-9
40. Sun S, Luo C, Chen J: A review of natural language processing techniques for opinion mining systems. Information Fusion. 2017, 36:10-25. 10.1016/j.inffus.2016.10.004
41. Almaraz-Damian J-A, Ponomaryov V, Sadovnychiy S, Castillejos-Fernandez H: Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. Entropy. 2020, 22:484. 10.3390/e22040484
42. Groh M, Badri O, Daneshjou R, et al.: Deep learning-aided decision support for diagnosis of skin disease across skin tones. Nature Medicine. 2024, 30:573-583. 10.1038/s41591-023-02728-3

43. Chiu M-C, Wang Y, Kuo Y-J, Chen P-Y: DDI-CoCo: A dataset for understanding the effect of color contrast in machine-assisted skin disease detection. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024. 6905-6909. 10.1109/ICASSP48485.2024.10448011
44. Ciążyńska M, Kamińska-Winciorek G, Lange D, et al.: Author Correction: The incidence and clinical analysis of non-melanoma skin cancer. Scientific Reports. 2021, 11:15705. 10.1038/s41598-021-94435-7
45. Sadik R, Majumder A, Biswas AA, Ahammad B, Rahman MM: An in-depth analysis of convolutional neural network architectures with transfer learning for skin disease diagnosis. Healthcare Analytics. 2023, 3:100143. 10.1016/j.health.2023.100143
46. Naeem A, Anees T, Fiza M, Naqvi RA, Lee SW: SCDNet: a deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. Sensors. 2022, 22:5652. 10.3390/s22155652
47. Tan M, Le Q: Efficientnet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. 2019, 6105-6114.
48. Behara K, Bhero E, Agee JT: Grid-based structural and dimensional skin cancer classification with self-featured optimized explainable deep convolutional neural networks. International Journal of Molecular Sciences. 2024, 25:1546. 10.3390/ijms25031546
49. Li M, Yuan B: 2D-LDA: A statistical linear discriminant analysis for image matrix. Pattern Recognition Letters. 2005, 26:527-532. 10.1016/j.patrec.2004.09.007
50. Das K, Cockerell CJ, Patil A, Pietkiewicz P, Giulini M, Grabbe S, Goldust M: Machine learning and its application in skin cancer. International Journal of Environmental Research and Public Health. 2021, 18:13409. 10.3390/ijerph182413409
51. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009, 339:b2535. 10.1136/bmj.b2535
52. Adebiyi A, Abdalnabi N, Smith EH, Hirner J, Simoes EJ, Becevic M, Rao P: Accurate skin lesion classification using multimodal learning on the HAM10000 dataset. MedRxiv. 2024, 10.1101/2024.05.30.24308213
53. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N: MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert Systems with Applications. 2015, 42:6578-6585. 10.1016/j.eswa.2015.04.034
54. Guha SR, Rafizul Haque SM: Convolutional neural network based skin lesion analysis for classifying melanoma. 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh. 2019, 1-5. 10.1109/STI47673.2019.9067979
55. Thapar P, Rakhra M, Alsaadi M, Quraishi A, Deka A, Ramesh JVN: A hybrid Grasshopper optimization algorithm for skin lesion segmentation and melanoma classification using deep learning. Healthcare Analytics. 2024, 5:100326. 10.1016/j.health.2024.100326
56. Liu H, Dou Y, Wang K, Zou Y, Sen G, Liu X, Li H: A skin disease classification model based on multi scale combined efficient channel attention module. Scientific Reports. 2025, 15:6116. 10.1038/s41598-025-90418-0
57. Haenssle HA, Fink C, Schneiderbauer R, et al.: Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology. 2018, 29:1836-1842. 10.1093/annonc/mdy166
58. Wei L, Ding K, Hu H: Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE Access. 2020, 8:99633-99647. 10.1109/access.2020.2997710
59. Krizhevsky A, Sutskever I, Hinton GE: ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017, 60:84-90. 10.1145/3065386
60. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. 10.1109/CVPR.2016.90
61. McCourt C, Dolan O, Gormley G: Malignant melanoma: a pictorial review. The Ulster Medical Journal. 2014, 83:103.
62. Gairola AK, Kumar V, Sahoo AK, Diwakar M, Singh P, Garg D: Multi-feature fusion deep network for skin disease diagnosis. Multimedia Tools and Applications. 2025, 84:419-444. 10.1007/s11042-024-18958-7
63. Seeja R, Suresh A: Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). Asian Pacific Journal of Cancer Prevention. 2019, 20:1555-1561. 10.31557/apjcp.2019.20.5.1555
64. Dorj UO, Lee KK, Choi JY, Lee M: The skin cancer classification using deep convolutional neural network. Multimedia Tools and Applications. 2018, 77:9909-9924. 10.1007/s11042-018-5714-1
65. Liu L, Liang C, Xue Y, et al.: An intelligent diagnostic model for melasma based on deep learning and multimode image input. Dermatology and Therapy. 2023, 13:569-579. 10.1007/s13555-022-00874-z
66. Naeem A, Anees T: A multiclassification framework for skin cancer detection by the concatenation of Xception and ResNet101. Journal of Computing & Biomedical Informatics. 2024, 6:205-227. 10.56979/602/2024
67. Li Y, Shen L: Skin lesion analysis towards melanoma detection using deep learning network. Sensors. 2018, 18:556. 10.3390/s18020556
68. Yu L, Chen H, Dou Q, Qin J, Heng PA: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging. 2017, 36:994-1004. 10.1109/tmi.2016.2642839
69. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Navab N, Hornegger J, Wells W, Frangi A (ed): Springer, Cham; 2015. 9351:234-241. 10.1007/978-3-319-24574-4_28
70. Naseri H, Safaei AA: Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review. BMC Cancer. 2025, 25:75. 10.1186/s12885-024-13423-y
71. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. arXiv. 2014, 10.48550/arXiv.1409.1556
72. Houssein EH, Abdelkareem DA, Hu G, Hameed MA, Ibrahim IA, Younan M: An effective multiclass skin cancer classification approach based on deep convolutional neural network. Cluster Computing. 2024, 27:12799-12819. 10.1007/s10586-024-04540-1
73. Bechelli S, Delhommelle J: Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. Bioengineering. 2022, 9:97. 10.3390/bioengineering9030097
74. Alsahafi YS, Kassem MA, Hosny KM: Skin-Net: a novel deep residual network for skin lesions classification

using multilevel feature extraction and cross-channel correlation with detection of outlier. Journal of Big Data. 2023, 10:105. 10.1186/s40537-023-00769-6

75. Słowińska M, Czarnecka I, Czarnecki R, et al.: Characteristics of patients with melanoma with non-melanoma skin cancer comorbidity: Practical implications based on a retrospective study. Oncology Letters. 2025, 29:214. 10.3892/ol.2025.14960
76. Purni JT, Vedhapriyavadhana R: EOSA-Net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks. Journal of King Saud University - Computer and Information Sciences. 2024, 36:102007. 10.1016/j.jksuci.2024.102007
77. Unnisa Z, Tariq A, Sarwar N, Din I, Serhani MA, Trabelsi Z: Impact of fine-tuning parameters of convolutional neural network for skin cancer detection. Scientific Reports. 2025, 15:14779. 10.1038/s41598-025-99529-0
78. Chollet F: Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, 1251-1258.