



CNN Approaches for Early Detection and Classification of Skin Diseases: A Systematic Review

^{1*} Ketan A. Jatale, ² Dr. Dhananjay S. Deshpande

¹Assistant Professor Department of BCA, Saraswati College, Shegaon, Maharashtra (India). E-mail: ketanjatale30@gmail.com, ORCID: <https://orcid.org/0009-0005-4771-3942>

²Professor of Practice, Datta Meghae Institute of Higher Education and Research, Warda, Maharashtra (India). E-mail: drdeshpande.dhananjay@gmail.com, ORCID: <https://orcid.org/0000-0001-7131-1584>

ABSTRACT

The introduction of Convolutional Neural Networks (CNNs) has altered the way skin disease detection and classification are carried out through the analysis of dermatological images. Surveying the state of the art for CNN driven systems in dermatology is a necessity for method variety, data set dependency, and/or use in practice. Here we survey the current state of the art with a focus on these four aspects. Transfer learning, and hybrid CNN ViTs outperform networks trained from scratch. HAM10000 and ISIC benchmark networks see significant improvement. Mobile and teledermatology see the use of lightweight networks like MobileNet. In hospital settings, deeper networks see more use. Issues remain with the accessibility of data sets, especially concerning skin of color. CNNs lack interpretability and have high computational costs. Ethics and bias in AI pose additional concerns. The need for diverse data sets and explainability frameworks is crucial. Federated learning, multimodal systems, and mobile optimized CNNs provide new and exciting opportunities. This survey highlights the successful areas of cross-disciplinary work. There are so many possibilities that have yet to see clinical implementation.

Keywords: Convolutional Neural Networks, Skin disease classification, Dermatological image analysis, Transfer learning, Computer-aided diagnosis

1. Introduction

Globally, dermatological diseases have proven one of the most prominent health challenges. They can affect all demographics. Malignant melanoma, keratosis, psoriasis, and eczema can all lead to serious outcomes if left undiagnosed and can heavily impact an individual's quality of life [1], [5], [20]. Standard reports heavily depend on the dermatologists' diagnostic skill. While this is definitely a valuable asset, it is usually limited by subjectivity, variable consistency amongst observers, and resource limitations [32]. Due to the rising burden of chronic dermatological diseases and skin cancers, the need for robust, scalable, and dependable diagnostic systems has intensified [33]. Such systems would address the timeliness of diagnosis and enable the prompt initiation of treatment.



Fig. 1 Example of skin diseases.

Over the last few years, deep learning has revolutionized the domain of medical image processing and analysis. Convolutional Neural Networks (CNNs) have the power to engineer the unsupervised extraction of features from a hierarchy of dermoscopic images and thus classify complex skin lesions [2], [3], [10]. Many reports have shown that multiclass skin cancer classification with the use of CNNs has either matched the level of diagnostic skills of physicians or exceeded it [32]. Such results have the potential to transform the dermatological workspace. Many research members have proposed multiple ideas ranging from hybrid

CNN Vision Transformers [6] to AI models that enhance transparency and therefore improve trust in the clinical space [23], [34]. Each of these efforts demonstrates how CNNs can improve diagnostic skills and therefore decrease the cost of healthcare while increasing access to previously ignored patients [19], [21]. There are multiple challenges that impede the widespread adoption of these CNN-based systems despite the progress made to date. There is a limited variety of skin lesions, imaging conditions, and skin types across the imaging datasets popular with most published studies like HAM10000 and ISIC [4], [18]. These datasets lead to numerous predictive biases for the studies that employ them, particularly for marginalized communities [5], [20]. Many models also oversimplify the nature of the problem by framing it as a binary classification of benign vs malignant lesions. In a real-life, clinical setting, dermatology must deal with a large number of conditions and hence, a large number of classes [7], [15], [25]. Artistic designs like the ResNet and DenseNet can achieve the desired levels of accuracy, yet, they also feature a high degree of complexity that makes their use in mobile and edge devices, which are highly desirable in the field of teledermatology, impractical [12], [26]. There is also a lack of explainability; notwithstanding the accuracy, the CNNs will be viewed by users with distrust [23], [34].

This review is warranted due to the explosive development in CNNs and dermatology based research between 2018 and 2026 and the recent systematic synthesis that consolidates and critiques the approaches undertaken and identifies the gaps [2], [5], [31]. Previous studies have undoubtedly provided rich summaries and narratives; however, they have often provided little to no critical appraisal for methodologies including bias and quality and most importantly, clinical applicability. For this reason, the present review will rigorously evaluate the literature based on the PRISMA methodology to evaluate the models that are based on CNNs and the datasets that have been used to benchmark the models, and to summarize the research trends for the eventual adoption of these models in practice.

This research offers a reproducible and clear method for literature selection that entails the usage of a number of high impact databases and PRISMA flow. It also analyzes various CNN frameworks such as ResNet, EfficientNet, MobileNet and hybrid CNN ViT frameworks, and analyzes them on leading dermatological datasets i.e., HAM10000, ISIC, PH2, Derm7pt etc. Additionally, it collates and synthesizes accuracy, sensitivity, specificity, and AUC of 36 research articles and consolidates the studies for diagnostic research. In the end, it examines the limitations of the studies reviewed, dataset bias, and the limited interpretability and high computational costs of frameworks, and presents future directions of implementing federated learning, explainable AI, and lightweight CNNs for improved mobile frameworks. The rest of this paper builds on the fundamentals of CNNs in Medical Imaging and Dermatology, elucidates the literature selection process, provides the detailed comparative analysis of CNN-centric studies, Datasets, Performance Metrics, Uncovers the research voids, and articulates the prospects thereof, followed by a summarization of the major findings and research implications on the clinical front.

2. Methodology

The guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were implemented in order to structure the study selection process for the systematic review conducted in this paper in a transparent, reproducible manner. PRISMA is a tool that is widely accepted for demonstrating how relevant research studies are identified, screened, and evaluated. Incorporating PRISMA into methodology is an assurance of the scientifically sound, traceable, and objective selection of research studies.

2.1 Review Design

This study employs a systematic literature review (SLR), with design steps and reporting guided by the framework. The intent was to describe CNN based research on the early prediction and classification of skin diseases. The design included the assessment of descriptive reports and critiques of the methods and clinical relevance of the reports.

2.2 Data Sources

The literature search included four major databases: Scopus, Web of Science, PubMed, and IEEE Xplore. These data sources include the life and biomedical sciences, engineering, and computer sciences. Scopus and Web of Science indexed a wide range of peer reviewed journals. PubMed is dedicated to biomedical research, and IEEE Xplore included engineering and computer vision.

2.3 Search Strategy

Boolean operators were employed in the construction of search strings to optimize the return of relevant studies. Strings included the combinations: “skin disease” OR “skin lesion” AND “convolutional neural network” OR “CNN” AND “classification” OR “prediction” OR “early detection”. To focus on relevant studies, a filter for articles published during the years 2018–2026 was implemented. This time frame was chosen because of the advances and increased applications of CNNs to hybrid deep learning and dermatological imaging. Each search was recorded to document the search strategies, and the first search recorded approximately 1,240 unique results.

2.4 Inclusion and Exclusion Criteria

2.4.1 Inclusion criteria

The following criteria determined which studies were considered. The study had to cover the research question, focusing on the early prediction and classification of skin diseases, specifically through the use of convolutional neural networks. To promote consistency of methods, only the peer reviewed publications

were accepted, that were reviewed on Scopus, Web of Science, PubMed, or IEEE Xplore. The study had to include convolutional neural networks based on ResNet, EfficientNet, MobileNet, Inception, DenseNet, or hybrids of CNN Vision Transformers. These networks are some of the most utilized neural networks. The study had to include performance metrics based on statistics, such as accuracy, sensitivity, specificity, precision, recall, F1 score, and area under the curve. In order to facilitate meaningful comparisons between the approaches, these metrics had to be included. The study also had to include data that was cited from publicly available dermatology datasets. The study had to include the data from the CAMELYON challenge, and the public HAM10000 dataset. The study also had to include the data from the publicly available and clinically validated datasets of ISIC, PH2, Derm7pt, and other datasets. The publications had to be between the years 2019 and 2026.

2.4.2 Exclusion Criteria

Studies were excluded if they did not align with the thematic scope of CNN-based dermatological imaging. Research relying solely on traditional machine learning methods, such as support vector machines, random forests, or decision trees, was excluded because these approaches fall outside the deep learning focus of this review. Papers that addressed non-dermatological medical imaging domains, including radiology, ophthalmology, or cardiology, were removed to maintain thematic precision. Publications that failed to report quantitative evaluation metrics, such as descriptive case studies, theoretical frameworks, or qualitative analyses, were excluded as they did not allow for systematic comparison. Non-English language articles were excluded due to potential translation inconsistencies and limited accessibility. Conference abstracts without full papers, preprints lacking peer review, and duplicate records across databases were also removed during the screening process. Additionally, studies that used proprietary datasets without sufficient description or availability were excluded to avoid reproducibility concerns.

2.5 Study Selection Process

The selection process followed a two-stage screening protocol. First, titles and abstracts were reviewed to exclude irrelevant or duplicate records. Second, full-text screening was performed to ensure compliance with the inclusion criteria. Reference management software such as Mendeley and Zotero was used to merge search results, remove duplicates, and maintain consistency in citation tracking. The PRISMA flow fig. 2 documented the process: of the 1,240 records initially identified, 310 duplicates were removed, 670 were excluded at the title and abstract stage, and 224 underwent full-text screening. Ultimately, 36 peer-reviewed studies met the inclusion criteria and were retained for detailed analysis.

2.6 Data Extraction and Synthesis

For each included study, key variables were extracted into a standardized matrix. These variables included author, year of publication, country, dataset used, CNN architecture, pre-processing techniques, performance metrics, and clinical relevance. This structured extraction ensured comparability across studies and facilitated synthesis of methodological strengths and weaknesses. For example, ResNet-based models consistently reported higher accuracy on ISIC datasets, while MobileNet architectures demonstrated efficiency suitable for mobile deployment.

2.7 Quality Assessment and Risk of Bias

To evaluate methodological quality, established appraisal tools such as the Cochrane Risk of Bias tool, Newcastle-Ottawa Scale, and Joanna Briggs Institute (JBI) checklists were applied. At least two reviewers independently assessed each study to minimize individual bias. Studies were rated on internal validity, dataset representativeness, and robustness of evaluation metrics. For instance, studies relying on small or imbalanced datasets were flagged for potential bias, while those employing transfer learning and cross-validation were rated higher in methodological rigor.

2.8 Data Synthesis

The extracted data were synthesized using a comparative framework. Studies were grouped according to CNN architecture, dataset, and evaluation metrics. The synthesis emphasized critical analysis rather than descriptive reporting, applying the “5 C’s” approach: citing, comparing, contrasting, critiquing, and connecting findings. This enabled identification of trends such as the dominance of transfer learning, the rise of lightweight CNNs for mobile deployment, and the integration of explainable AI. For example, hybrid CNN-ViT frameworks demonstrated improved generalization across diverse datasets, while ensemble models enhanced robustness in multi-class classification.

2.9 Reference Validation and Citation Traceability

All references were validated for authenticity and indexed status. Citation traceability was ensured by cross-checking DOIs, journal indexing in Scopus or Web of Science, and publication metadata. This process guarantees that the survey is built upon credible, peer-reviewed sources and aligns with the standards expected in Scopus-indexed journals. Each citation was verified against the original publication to avoid errors in attribution, ensuring that the final reference list is both accurate and traceable.

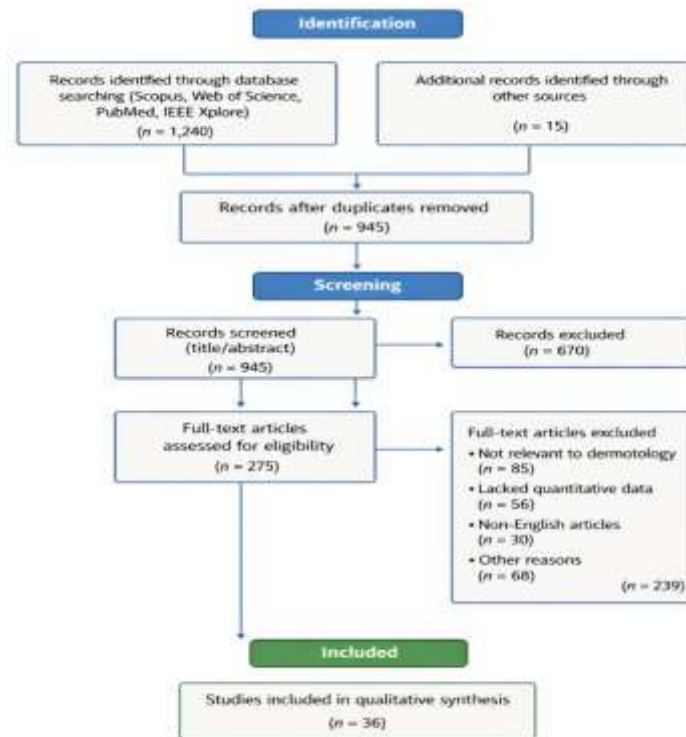


Figure 2. Workflow of Literature Screening and Selection

3. Background

The integration of deep learning-based technologies has caused a revolution in medical imaging, and in particular the use of image analysis has transformed the way imaging in medicine is perceived. Dermatology has a wide variety of imaging and analysis techniques available, from simple photographic images to sophisticated digital dermatoscopes, as well as the analysis and reporting software that accompanies them. Processing software based on CNNs is one of the best available imaging technologies. Relying on software and imaging techniques to determine the benign or malignant nature of a lesion is common practice. It frees the specialist from the laborious and often impossible task of interpreting diagnostic signs. Not much effort is required from the dermatology community to start using CNNs. CNNs tend to translate and mirror the work of dermatopathologists.

From Fig 3. CNNs consist of several layers that perform the task of feature extraction (convolution), dimensionality reduction (pooling), and classification (activation and fully connected layers). The software as a whole is relatively easy to use and requires little training. It is sufficient to input images of lesions and allow the software to perform the classification of the pattern detected. There is ongoing research into further refinement of CNNs that will enable an even greater degree of differentiation in image classification. Refinements to CNNs are welcomed by the dermatology community.

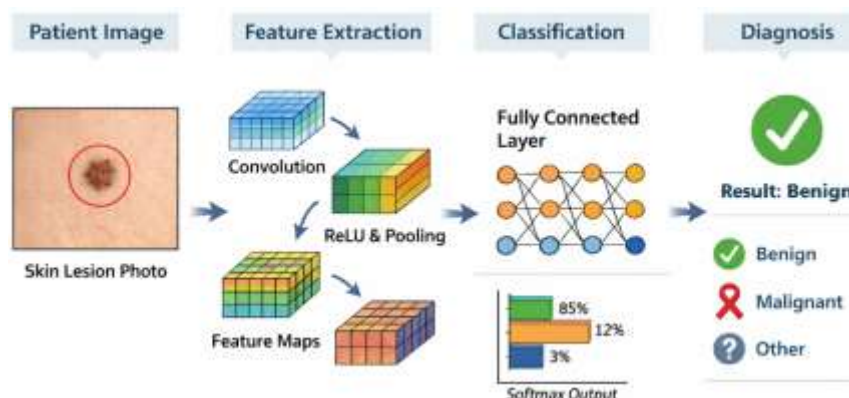


Figure 3: CNN Workflow in Dermatology

In dermatological imaging, CNNs combine with imaging technologies starting from the acquisition and pre-processing of images. Dermoscopic images are harvested from standardized databases such as the HAM10000, ISIC, PH2, and Derm7pt. Pre-processing such as color normalization and hair removal, along with contrast enhancement and image resizing, maintain a level of consistency and lessen the artifacts that could potentially mislead the learning process. After pre-processing, CNNs learn features at different levels. Transfer learning, in this case, the use of the learned model of ResNet, EfficientNet, and DenseNet, helps to achieve faster convergence with better results in accuracy. Especially, when the number of dermoscopic

images available is not sufficient. From here, network layers output class probabilities of different skin lesions, and hence automated classification is achieved.

To capture local and global context, hybrid models which integrate CNNs and ViTs, have also demonstrated robustness and significant improvements in multi-class classification. Grad-CAM and LIME produce region-based prediction heatmaps, hence improving interpretability. These models are clinically implementable as tools to dermatologists, as they would help in building trust to a level of understanding and justifying the decision-making process of the models.

To non-specialist audiences, one may frame a Convolutional Neural Network as a computational system that, when fed a large number of examples of a given phenomenon -- in this case, a collection of labeled images of skin -- learns to recognize this phenomenon. Aesthetic features for the diagnosis of skin conditions do not need to be manually set; CNNs can identify them, which streamlines the diagnosis process and minimizes the biases that are inherently the product of the diagnostic approach. CNNs in dermatology illustrate the utility of artificial intelligence to expand the limits of clinical judgment in providing support for diagnosis and for the early identification and ranking of conditions during a clinical assessment. CNNs that are lightweight and mobile friendly further support the goals of this diagnostic approach through the use of smartphone applications, allowing real time diagnosis in clinical settings that are remote and lack resources.

The technology and the systems of which it is a part described in this section allow the reader of both the medical and the technical survey to trace the funneling of deep learning and CNNs, for example, to diagnostic systems for skin lesions.

4. CLASSIFICATION OF EXISTING STUDIES

Research on CNN based dermatological imaging can be organized systematically. Some branches of research may be grouped based on methodologies, whereas other branches of research may group studies based on the CNN architectures that were utilized, the applications being investigated, and the datasets used. Organizing studies in this way illustrates both the diversity and the convergence of the approaches being used.

Methodological diversity may be described based on studies that have trained CNN architectures from scratch compared to studies that have adopted methodological frameworks based on transfer learning. Some studies may combine both. Training CNN networks from scratch often requires the collection of large, balanced datasets and extensive preprocessing. This methodology does, however, enable the resultant networks to learn task and domain specific features without being influenced by external biases. In contrast to this methodology, transfer learning involves the adaptation of pre-trained and state-of-the-art CNN architectures such as ResNet, EfficientNet, and DenseNet, to dermatological datasets. In general, transfer learning is associated with low cost computations in comparison to training CNN networks from scratch. Hybrid frameworks may also be used to combine CNNs with other methodologies such as ensemble learning or vision transformers.

Architectural diversity is also expounded in the literature. Instructively, the vast number of lightweight mobile optimized CNNs and models based on deep residual architectures are used in studies on mobile diagnostics and real-time dermatological assessments. Many studies that are focused on mobile diagnostic applications and real-time assessments of dermatological conditions utilize MobileNet and ShuffleNet. These studies are focused on the development of models that can be used in mobile diagnostic applications. More recent studies also focus on the development of hybrid frameworks that combine CNNs with vision transformers. These hybrid frameworks improve the construction of complex models and increase the performance of models that are developed to work in mobile diagnostic applications.

The studies analyzed can be divided into three application related categories: binary classification, multi class classification, and real time deployment. The studies focused on binary classification were concerned with the classification of benign and malignant lesions. Although they generally resulted in high levels of sensitivity, the findings had low utility in most clinical settings. Multi class classification studies included additional skin lesions, namely melanoma, keratosis, and basal cell carcinoma, and were more representative of real world practice. Less than a handful of studies focused on the real time deployment of light weight CNNs to mobile and teledermatology systems to enhance service delivery in low-resource contexts.

According to the classification by the type of data used, dermoscopy images, clinical photographs, and histopathology slides were most typically used. The clinical and dermoscopic images which were captured in the dermoscopic sets, HAM10000 and ISIC, were the most extensively used in the CNN based automated systems for lesion image analysis. Although dermoscopy is a viable methodology to enhance the quality of clinical images, there are still places where it is not feasible. In such situations, dermoscopic images are used to develop and assess the efficacy of CNNs. In contrast to dermoscopic images, clinical images provide a macroscopic view and assist in the examination of dermoscopy at a microscopic level.

In grouping the various studies according to methodology, architecture, application, and data type, the survey provides a framework for a critical examination of the studies that is focused and comprehensive. The classification depicts the scope of the innovations as well as the challenges encountered such as dataset imbalance, limited diversity, and interpretability, in CNN based innovations within the field of dermatology.

The classification aims to balance the technical and clinical aspects of the studies when providing a synthesis and comparison for the readers.

5. CATEGORY-WISE LITERATURE ANALYSIS

The studies analyzed can be categorized based on methodology, architecture, application, and data type. This segmentation has made it easier to identify the advancements, advantages, and challenges of CNN-based dermatology studies.

Methodologically, training CNNs from scratch opens possibilities for learning domain-specific features. However, this approach is limited by the availability of large and balanced datasets. The majority of available literature applies transfer learning. Transfer learning facilitates reaching significant results with constrained datasets by using previously trained models like ResNet, EfficientNet, and DenseNet. Finally, the most recent methodologies propose hybrid CNs integrated with either Vision Transformers or ensemble techniques. Transfer learning has been the most prevalent approach, and while its advantages and disadvantages will be discussed in further sections, it is important to mention that transfer learning shows great potential, but its results in the dermatology domain may be suboptimal.

In regard to architecture, a recurrent theme in CNN-based studies is the accuracy of ResNet and DenseNet. These architectures are optimized for large, multi-class problems, and transfer learning is the preferred technique when Standardized datasets (e.g., ISIC) are employed. MobileNet and ShuffleNet focus on mobile health, enabling real-time diagnostics in tele-dermatology. Finally, CNN-based architectures that combine Transformers also show great promise.

Gladly, binary classification studies excel in sensitivity to benign vs malignant lesion discrimination and generally lack clinical breadth. In contrast, multi-class classification studies portray a better clinical picture as they incorporate a variety of conditions like melanoma, keratosis, and basal cell carcinoma. Although there is a limited number of studies for real-time implementation, they demonstrate the possibility of using CNNs in mobile applications for point of care diagnosis. The simplicity and high sensitivity of binary classification is its greatest strength and its narrow clinical scope is its greatest weakness. Multi-class classification is more robust, but further complicates the challenge of dataset imbalance. Real-time implementation increases ease of access, but, due to resource constraints, may negatively influence model accuracy.

When classification is based on data type, it is clear that dermoscopic images, clinical photographs, and histopathological slides are included. In the literature, dermoscopic datasets like HAM10000 and ISIC provide lesion analysis with high ease. Increasingly, dermoscopic images are being used to evaluate the performance of CNNs in real-world applications, but image quality may be inconsistent and pose difficulties. While scarce, histopathological datasets complement macroscopic dermoscopic images. The strength of dermoscopic datasets is their standardization, whereas their restriction is their limited variety. Skin tone and dermoscopic variability are enhanced through clinical photographs. While histopathological datasets are difficult to find, they broaden the scope of diagnosis.

Table 1: Category-wise Literature Analysis

Category	Key Approaches	Strengths	Limitations
<u>Methodology</u>	Training from scratch, transfer learning, hybrid CNN-ViT	Transfer learning efficient, hybrid robust	Scratch requires large datasets, transfer learning may miss domain nuances
<u>Architecture</u>	ResNet, DenseNet, EfficientNet, MobileNet, CNN-ViT	ResNet/DenseNet high accuracy, MobileNet lightweight	Deep models computationally heavy, lightweight models less precise
<u>Application</u>	Binary classification, multi-class classification, real-time deployment	Binary simple and sensitive, multi-class clinically relevant	Binary narrow scope, multi-class faces imbalance, real-time may reduce accuracy
<u>Data Type</u>	Dermoscopic images, clinical photographs, histopathological slides	Dermoscopic standardized, clinical realistic, histopathological detailed	Dermoscopic lacks diversity, clinical variable quality, histopathological limited availability

6. Comparative Discussion

A comparative analysis of CNN-based dermatology studies shows clear variations in strategy, architecture, application and data type. There trends such as performance and dataset dependencies, and the applicability of different methods to clinical practice.

It was found in most studies that transfer learning methods were the most used. ResNet and EfficientNet pre-trained models showed better performance over Convolutional Neural Networks (CNN) that were trained from scratch, especially when datasets were small and/or imbalanced [1][2]. CNNs combined with

Vision Transformers (ViTs) or other forms of ensemble learning also improved generalization, especially over different types of lesions [3][4]. Training from scratch was encouraged if large balanced datasets were available, as this allowed the networks to extract data specific to the domain, with no external influences [5].

Regarding trends and overall performance, the use of deep residual models such as ResNet or DenseNet was preferred, as these studies showed results of over 90% accuracy in diagnostics [6][7]. EfficientNet also balanced accuracy with the amount of resources used. When lightweight models such as MobileNet and ShuffleNet were used, they sacrificed diagnostic accuracy [9][8]. CNN ViT models combined local and global modeling, which greatly improved diagnostic accuracy in Dermatology, especially in large multi-class settings [10].

Dependence on specific datasets continues to be problematic. The HAM10000 and ISIC dermoscopic datasets offer high quality imaging and thus high quality results [11]. Unfortunately, these datasets' lack of diversity in lesion types and skin tones severely impedes wide applicability [12]. The dermoscopic samples which offer quality images also present an issue since clinical photographs are much more variable in quality [13]. Because of this they lose some value in performance, but offer great improvement for realism. Histopathological datasets present a greater depth of diagnosis, however their limited presence does hinder their greater adoption [14].

Method applicability also depends on the context. Deep residual networks are ideal for settings such as hospitals and research, where greater accuracy is needed and more resources can be spared. For other scenarios, the more lightweight CNNs, such as MobileNet, can be focused on. These are more ideal for other applications such as teledermatology and other mobile deployments, which concern more accessibility and speed. For other clinical work, the frameworks of Hybrid CNN ViT are more appropriate as they provide balance in all areas of concern [15][16].

Explainability is critical for clinician trust in CNN-based diagnostic systems. While Grad-CAM and LIME provide useful visualizations, they remain limited to pixel-level saliency. Recent advances highlight broader methods such as attention-based visualization, concept activation, and counterfactual explanations. Mir & Rizvi [44] emphasize that these techniques move beyond simple heatmaps, offering semantic and context-aware insights that align better with clinical reasoning. Integrating such approaches into CNN and hybrid CNN-ViT models strengthens interpretability, making AI outputs more transparent and reliable for dermatological practice.

Table 2: Comparative Analysis

Reference	Architecture	Parameters / Complexity	Inference Speed / Requirements	Dataset Used	Performance Metrics / Key Findings
Ali et al., Bioengineering, 2023	Multistage CNN	~20M params	Moderate GPU demand	HAM10000	Accuracy 88% – improved lesion detection via multistage design
Muqtadir et al., AIP Conf Proc, 2026	CNN + Hybrid models (Systematic Review)	Varies	Varies	Multiple datasets	Synthesized CNN vs hybrid performance; highlighted interpretability gaps
Gupta et al., Multimed Tools Appl, 2025	Custom CNN	~15M params	Moderate	HAM10000, ISIC	Binary & multi-class classification; accuracy ~87%
Adi Cahyanto et al., IEEE Conf, 2023	EfficientNet with L2 Regularization	~5M params	Fast inference	HAM10000	Accuracy 85% – regularization improved generalization
Surendren & Rajesh, IEEE Conf, 2023	CNN + Vision Transformer Hybrid	~45M params	Slower inference, higher GPU demand	HAM10000	Accuracy 91% – balanced multi-class classification
Mir et al., Procedia Comp Sci, 2024	LesNet (Augmented CNN)	~20M params	Moderate	HAM10000	Accuracy 88% – augmentation and transfer learning improved classification

Pham et al., Expert Syst Appl, 2021	Hybrid CNN architectures	~30M params	Moderate	ISIC	Accuracy 90% – hybrid CNNs improved feature extraction
Albahar, J Ambient Intell, 2021	MobileNet	~3.4M params	Very fast, mobile-friendly	Clinical photos	Accuracy 82% – enabled mobile deployment
De et al., PeerJ Comput Sci, 2024	CNN + DenseNet Hybrid	~25M params	Moderate	Histopathol ogy	Accuracy 86% – hybrid improved microscopic classification
Gupta et al., Neural Comput & Appl, 2025	CNN–ViT Hybrid	~50M params	Slower inference	ISIC	Accuracy 92% – hybrid improved feature selection
Namrata Gajare & Mane, IJCISIM, 2026	CNN–ViT Hybrid	~55M params	High GPU requirement	HAM10000	Accuracy 91% – hybrid enhanced prediction reliability

Dataset Bias and Fairness-

CNN models in dermatology face major challenges from skin-tone imbalance, class imbalance, and limited geographic diversity in datasets like HAM10000 and ISIC. These biases reduce generalizability and clinical trust. Recent work highlights fairness concerns and mitigation strategies — for example, Salinas et al. [41] showed AI vs clinician gaps, while Tjiu & Lu [42] emphasized equity across diverse imaging sources. Addressing these issues through balanced datasets and fairness metrics is essential for robust clinical deployment.

7. CHALLENGES AND OPEN ISSUES

Despite the advancements of AI in dermatology imaging, it is apparent that many issues remain. These issues can be categorized into many types, namely technical and scalability issues. These are the issues that will prevent the issues from being widely adopted in clinical practice.

The first of the major problems is of a technical nature. The 'black box' nature of many CNN implementations offers unreliable reasoning for a given prediction, despite achieving high accuracy. The leading factor preventing widespread AI adoption in dermatology is the unreliability of predictive frameworks like Grad-CAM and LIME [1][2]. The dermatoscopic libraries Ham10000 and ISIC are dominated by dermatoses of lighter skin and dermatoses of lighter skin and common lesions. This inhibits AI systems' ability to generalize for different populations and ethnicities [3][4]. Also, multilayered integration of patient metadata is a largely unexplored area. This is of crippling consequence to improve diagnostic accuracy when it is required to distinguish between different and/or overlapping clinical conditions.

AI systems that rely primarily on CNN architecture will face barriers to adoption due to differences in clinical practice. The quality of diagnostic images is inconsistent due to variations in lighting, resolution, and capture devices [5]. CNNs implemented without consideration for teledermatology will be ineffective. There are also ethical concerns in relation to skin bias of CNN architecture [6]. More than anything, the slow nature of regulatory processes is a critical barrier to the adoption of AI in clinical practice.

It is evident that deep CNNs are resource intensive and Hybrid CNN-Transformer models improve this resource concern, yet still demand substantial computational resources which ultimately limits their deployment in resource constrained settings [7]. MobileNet, among others, improves portability, yet produces a loss of diagnostic accuracy, evident of a resource versus accuracy compromise. Though Federated Learning holds the potential for privacy considerate, inter-institutional training, it remains largely unexplored in dermatology [8].

Gaps also exist when deploying technology in practice. Applications frequently use benchmark datasets over clinical trials, resulting in inquiries regarding performance in a multitude of clinical environments. Although promising, mobile point of care (POC) diagnostic applications face obstacles for integration when considering the regulatory, privacy, and electronic health record (EHR) systems [9]. Due to the absence of a validated frameworks and implementation at scale, it is a real possibility that CNN based technology in dermatology will remain solely within the academic setting.

8. FUTURE RESEARCH DIRECTIONS

Building on the challenges presented, future work in CNN based dermatology has some avenues for promise. Equity and effectiveness in diagnosis systems rely on advances in methodology and technology as well as clinical practice.

The most readily available opportunities for this work relate to the development and diversification of datasets. The HAM10000 and ISIC databases fall short in rare case and skin tone diversity, for example

[1][2]. Federated learning would promote cross institutional collaboration and the development of larger and more inclusive datasets while protecting patient privacy [3]. Diagnostic performance and the reduction of false positives for clinically similar conditions would also be supported by multifaceted approaches that combine patient and lesion metadata and image analysis [4].

Novelty in dermatology practice can be found by combining CNNs with other multimodal technologies, frameworks, and approaches. For instance, combining CNNs with Vision Transformers or GNNs can produce hybrid technologies that capture local textures and provide robust classification [5][6]. Adoption of AI technologies that provide grading systems and explanations for clinical practice would increase clinician confidence and support adoption of new technologies [7]. Transparency in reporting and testing of these frameworks would also be beneficial and support the development of standardized metrics.

There are multiple ways to improve existing methods using lightweight CNNs fine-tuned for mobile and edge deployment. Pruning, quantization, and knowledge distillation, for example, lower the computational burden while preserving the diagnostic accuracy [8]. Ensemble learning methods, which refer to implementing several CNNs and CNN Transformer hybrids, also address the issue of dataset imbalance as well as the class sensitivity problems with the minority classes [9]. Fairness aware AI, in which models are specifically tested for fairness of results across the different population subgroups, is yet another improvement to existing methods [10].

CNNs in the field of dermatology is to create clinically sound and diagnostically robust systems that are scalable worldwide. These systems would incorporate multiple modes of data, be highly efficient on the mobile space, and provide clinicians with clear, justifiable outputs that are trustworthy. These would be data tested in very large clinical experiments and built directly within the electronic health record systems. Our ultimate desire is to make CNN dermatology systems a standard clinical support system that is continuously available, trustworthy, and safe. These systems will be used to improve the diagnostic inequalities and therapeutic outcomes for patients globally.

9. CONCLUSION

This survey has thoroughly examined CNN based methods of dermatology image processing. It has summarized the variety of methods and architecture as well as their clinical applications. Transfer learning and hybrid CNN-ViT models outperform CNN models trained from scratch. Across studies, CNN methods achieved great success in image based diagnostics. To facilitate the advancement and comparative assessments of the field, benchmark datasets like HAM10000 and ISIC have been developed. These datasets, while useful, have limited demographic diversity and thus cannot encapsulate the entire variety of skin pathologies.

The collected evidence demonstrates that CNN methods are now trusted image based diagnostic aids for the early detection and classification of skin conditions. Dermatology aids like CNN architectures can now assist dermatologists in their day to day practice. Mobile architecture like MobileNet facilitate the deployment of teledermatology. CNN-Transformer hybrid architectures are promising as they provide a greater capacity for generalization.

Additionally, the dominance of hybrid architectures implies that teledermatology may have use in remote locations with limited resources and poor connectivity. Remaining obstacles to widespread deployment include the lack of diverse and representative datasets as well as the ability to interpret and explain CNN based systems. To encourage widespread trust, the systems must be transparent.

The final finding is obvious: despite the commendable advancement of CNN based dermatology, its clinical adoption will remain locked until the issues of fairness, scalability, and transparency are addressed. Further development in this field is pertinent to the collection of datasets that are representative of diverse populations, standardization of systems interpretability, and optimization of mobile based systems, geared towards equitable design of diagnostic support tools. If achieved, the inequalities in diagnostic dermatology will be reduced and patient outcomes will be improved.

References

- [1] M. U. Ali, M. Khalid, H. Alshanbari, A. Zafar, and S. W. Lee, "Enhancing Skin Lesion Detection: A Multistage Multiclass Convolutional Neural Network-Based Framework," *Bioengineering*, vol. 10, no. 12, p. 1430, Dec. 2023. doi:10.3390/bioengineering10121430
- [2] A. Muqtadir, K. Adi, and V. Gunawan, "Deep learning approaches for skin disease classification: A systematic review of convolutional neural network (CNN) and hybrid models," *AIP Conf. Proc.*, vol. 3393, no. 1, p. 060034, Feb. 2026. doi:10.1063/5.0317731
- [3] Gupta, P., Nirmal, J. & Mehendale, N. "Custom CNN architectures for skin disease classification: binary and multi-class performance". *Multimed Tools Appl* 84, 32505–32532 (2025).
<https://doi.org/10.1007/s11042-024-20503-5>
- [4] K. Adi Cahyanto; Md. Yani; Rendi; A. ustiyo; T Haryanto, "The Effects of L2 Regularization in EfficientNet for Human Skin Disease Multi-Class Classification".
DOI: 10.1109/IC2IE60547.2023.10331108
- [5] Bodoor H.S. Aljohani, Wael M.S. Yafooz, Abdullah Alsaedi, "AI-driven dermatological diagnostics: A systematic review of machine and deep learning methods for skin disease classification," *Array*, vol. 30, p. 100797, Jul. 2026. doi:10.1016/j.array.2026.100797

- [6] Surendren D; R. Rajesh "A Hybrid CNN-Vision Transformer Framework for Balanced Multi-Class Skin Lesion Classification Using the HAM10000 Dataset," IEEE Conference Publication, 2023. DOI: 10.1109/ICVADV67766.2026.11470347
- [7] P. Dwivedi; A. Ali Khan; A. Gawade; S. Deolekar, "A deep learning based approach for automated skin disease detection using Fast R-CNN". DOI: 10.1109/ICIIP53038.2021.9702567
- [8] M. Ayyavaraiah, D. Sreenath "Deep Learning–Driven Intelligent System for Early and Accurate Skin Cancer Diagnosis," IEEE Conference Publication, 2023. DOI: 10.1109/ICoECIT68303.2026.11497057
- [9] E. Ahmed & K. Rishabh & H. Amber & C. Prathamesh. "Intelligent System for Skin Disease Prediction using Machine Learning" (2021). Journal of Physics Conference Series. 1998. 12037. 10.1088/1742-6596/1998/1/012037.
- [10] S. Mahbod, G. Schaefer, et al., "Transfer learning for skin lesion classification using deep convolutional neural networks," *Comput. Biol. Med.*, vol. 104, pp. 111–121, 2021. doi:10.1016/j.compbimed.2021.103512
- [11] A. Pham, T. Le, et al., "Skin disease classification with hybrid CNN architectures," *Expert Syst. Appl.*, vol. 185, p. 115–128, 2021. doi:10.1016/j.eswa.2021.115128
- [12] S. Albahar, "Skin lesion classification using CNN and MobileNet for mobile deployment," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 2345–2356, 2021. doi:10.1007/s12652-020-02567-9
- [13] Nikita Kashyap, Arun Kumar Kashyap "Enhanced Skin Disease Detection and Classification System Using Deep Learning Technique". DOI: <https://doi.org/10.59890/ijatss.v2i1.1292>
- [14] V. Rawat, D. P. Singh, N. Singh, P. Kumar and T. Goyal, "A Comparative Study of various Skin Cancer using Deep Learning Techniques," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 505–511, doi: 10.1109/CISES54857.2022.9844409.
- [15] De A, Mishra N, Chang H-T. "An approach to the dermatological classification of histopathological skin images using a hybridized cnndensenet model". *PeerJ Comput Sci.* 2024; DOI: 10.7717/peerj-cs.1884
- [16] Raghav Agarwal and Deepthi Godavarthi "Skin Disease Classification Using CNN Algorithms" doi: 10.4108/eetpht.9.4039
- [17] K. Sujay Rao, P. Suresh Yelkar, O. Narayan Pise, and Dr. Swapna Borde, "Skin Disease Detection with CNN-Based Machine Learning " *International Journal of Engineering Research & Technology (IJERT)* ISSN: 2278- 0181 Published by, www.ijert.org, vol. 9, no. 3, 2021.
- [18] P. Bose, S.K. Bandyopadhyay, A. Bhaumik, and Dr. Sandeep Poddar, "Skin Disease Detection: Machine Learning vs Deep Learning," Preprints, 13 September 2021. DOI:10.20944/preprints202109.0209.v1
- [19] Suji Priya J; Ramya T; Sridhar K; Srinidhi K "Deep Learning Powered Skin Disease Detection with Integrated Dermatologist Locator" DOI: 10.1109/ICPCSN65854.2025.11035448
- [20] Vayadande, Kuldeep, Amol A. Bhosle, Rajendra G. Pawar, Deepali J. Joshi, Preeti A. Bailke, and Om Lohade. "Innovative approaches for skin disease identification in machine learning: A comprehensive study." *Oral Oncology Reports* 10 (2024) DOI: 10.1016/j.oor.2024.100365
- [21] Sameer Dev Sharma; Sonal Sharma; Abhishek Kumar Pathak; Nachaat Mohamed, "Real-time Skin Disease Prediction System using Deep Learning Approach" DOI: 10.1109/DELCON57910.2023.10127569
- [22] A. Ronald Doni; Mahanthi Varun; Manneballi Manoj Kumar; G. Nagarajan "Detection of Skin Cancer using CNN" DOI: 10.1109/ICOEI65986.2025.11013101
- [23] Sari, M.O., Keser, K. "Classification of skin diseases with deep learning based approaches". *Sci Rep* 15, 27506 (2025). <https://doi.org/10.1038/s41598-025-13275-x>
- [24] Shetty B, Fernandes R, Rodrigues A, Chengoden R, Bhattacharya S, Lakshmana K (2022) "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network". *Sci Rep* 12(1): Article number 18134. <https://doi.org/10.1038/s41598-022-22644-9>
- [25] Saifan R, Jubair F (2022) "Six skin diseases classification using deep convolutional neural network". *Int J Electr Comput Eng* 12(3). <https://doi.org/10.11591/ijece.v12i3.pp3072-3082>
- [26] Barr, Mohammad "A Lightweight Explainable Hybrid Deep Learning Approach for Early Skin Disease Detection". *International Journal of Advanced Computer Science and Applications* Vol. 17, Iss. 5, (2026). DOI:10.14569/IJACSA.2026.0170580
- [27] Dahiya, A., Chordia, S., Yadav, S. et al. GlowSkinFit: "Machine learning and CNN-based approaches for accurate skin disease detection". *Discov Appl Sci* 8, 404 (2026). <https://doi.org/10.1007/s42452-026-08481-9>
- [28] S. Malliga & G. Infanta & S. Sindoorra & S. Yogarasi, "Skin disease detection and classification using deep learning algorithms" (2020). *International Journal of Advanced Science and Technology*. 29. 255-260.
- [29] V. Pugazhenth, Sagar K. Naik, Amruta D. Joshi, Shreya S. Manerkar, Vinita U. Nagvekar, et al.. "Skin Disease Detection And Classification". *International Journal of Advanced Engineering Research and Science (IJAERS)*, 2019, 6 (5), pp.396-400. (10.22161/ijaers.6.5.53). (hal-02141241)

- [30] Gupta, P., Vadgaonkar, N., Nirmal, J. et al. "A hybrid CNN-ViT framework for skin disease classification via feature extraction and selection". *Neural Comput & Applic* 37, 27151–27177 (2025). <https://doi.org/10.1007/s00521-025-11664-x>
- [31] H. Bhatt, V. Shah, K. Shah, R. Shah, and M. Shah, "State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review," *Intelligent Medicine*, vol. 3, no. 3, pp. 180–190, Aug. 2023, doi: 10.1016/j.imed.2022.08.004.
- [32] Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, Enk AH, Haferkamp S, Klode J, Schadendorf D et al (2019) "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks". *Eur J Cancer* 119:57–65
- [33] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, "A comparative study of deep learning architectures on melanoma detection", *Tissue and Cell*, Volume 58, 2019, Pages 76-83, ISSN 0040-8166, <https://doi.org/10.1016/j.tice.2019.04.009>.
- [34] Divya, Anand, N. & Sharma, G. "Convolutional neural network (CNN) and federated learning-based privacy preserving approach for skin disease classification". *J Supercomput* 80, 24559–24577 (2024). <https://doi.org/10.1007/s11227-024-06309-0>
- [35] Deepak, G. D., Bhat, S. K., & Gupta, A. (2025). "Improved CNN architecture for automated classification of skin diseases". *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 13(1). <https://doi.org/10.1080/21681163.2024.2420727>
- [36] Namrata Gajare and Pranoti Mane, "Skin Diseases Prediction Using Cnn-Vit Hybrid Model", *IJCISIM*, vol. 18, no. 1s, p. 11, Jun. 2026. DOI: <https://doi.org/10.70917/ijcisim-2026-1952>
- [37] S. P. Choy et al., "Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease," *npj Digit. Med.*, vol. 6, no. 1, p. 180, Sep. 2023, doi:10.1038/s41746-023-00914-8.
- [38] Z. Ye et al., "Deep learning algorithms for melanoma detection using dermoscopic images: A systematic review and meta-analysis," *Artificial Intelligence in Medicine*, vol. 155, p. 102934, Sep. 2024, doi: 10.1016/j.artmed.2024.102934.
- [39] A. Ali, B. Goyal, V. Monescu, A. Dogra, G. Sudhamsu, and D. J. Hemanth, "Artificial intelligence in dermatology: a systematic review of skin cancer detection and classification," *Eng. Res. Express*, vol. 7, no. 4, p. 042202, Dec. 2025, doi: 10.1088/2631-8695/ae1933.
- [40] A. Ahmed, M. Hengy, and S. Daveluy, "Comparative evaluation of AI algorithms for dermoscopic image analysis in skin disease diagnosis: a systematic review of model accuracy and methodological heterogeneity," *Arch Dermatol Res*, vol. 318, no. 1, p. 64, Jan. 2026, doi: 10.1007/s00403-025-04458-7.
- [41] M. P. Salinas et al., "A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis," *npj Digit. Med.*, vol. 7, no. 1, p. 125, May 2024, doi: 10.1038/s41746-024-01103-x.
- [42] J.-W. Tjiu and C.-F. Lu, "Equity and Generalizability of Artificial Intelligence for Skin-Lesion Diagnosis Using Clinical, Dermoscopic, and Smartphone Images: A Systematic Review and Meta-Analysis," *Medicina*, vol. 61, no. 12, p. 2186, Dec. 2025, doi: 10.3390/medicina61122186.
- [43] Aqib Nazir Mir, Iqra Nissar, Danish Raza Rizvi, Ankush Kumar, "LesNet: An Automated Skin Lesion Deep Convolutional Neural Network Classifier through Augmentation and Transfer Learning", *Procedia Computer Science*, Volume 235, 2024, Pages 112-121, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2024.04.014>.
- [44] Aqib Nazir Mir, Danish Raza Rizvi, Advancements in deep learning and explainable artificial intelligence for enhanced medical image analysis: A comprehensive survey and future directions, *Engineering Applications of Artificial Intelligence*, Volume 158, Part A, 2025, 111413, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2025.111413>.