iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76121

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Review of Recent Deep Neural Networks for Skin Cancer Classification

Naresh Kumar Dewangan[1], Ankita Singh Baghel[2], Abhishek Guru[3]
Department of Computer Science and Engineering, MATS School of Engineering & Information Technology, Raipur, Chhattisgarh, India

Abstract: Skin cancer remains one of the most prevalent and rapidly rising malignancies worldwide, emphasizing the need for accurate and early detection through automated diagnostic tools. Deep learning has significantly advanced dermoscopic image analysis, with Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) emerging as two dominant paradigms. CNNs excel at capturing fine-grained local texture patterns such as pigment networks, color variations, and border irregularities, while ViTs leverage self-attention mechanisms to model long-range global dependencies and holistic lesion structures. This review provides an in-depth examination of CNN-based and ViT-based skin cancer classifiers, discussing their architectural principles, feature extraction capabilities, performance trends, and suitability for real-world clinical settings. We analyze key publicly available skin cancer datasets, preprocessing pipelines, training strategies, and evaluation metrics commonly used with these models. Furthermore, we highlight the complementary strengths of CNNs and ViTs, assess recent hybrid architectures that integrate local and global feature learning, and discuss challenges related to data imbalance, domain variability, computation cost, and model interpretability. The review concludes by outlining future research opportunities toward developing robust, transparent, and clinically reliable AI systems using CNN, ViT, and hybrid approaches for skin cancer diagnosis.

Keywords: Skin cancer classification, Dermoscopic images, Convolutional Neural Networks (CNNs), Vision Transformers (ViTs)

I. INTRODUCTION

Skin cancer is one of the fastest-growing malignancies globally, with its incidence rising steadily across diverse age groups and geographical regions [1]. Millions of new cases are reported every year, making it a major public health concern. Although basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) constitute the majority of diagnoses, melanoma despite being less common is responsible for most skin cancer-related deaths due to its high metastatic potential and rapid progression. The global increase in ultraviolet (UV) exposure, lifestyle changes, genetic predisposition, and inadequate screening practices further contribute to the growing burden. Early detection remains the single most effective strategy to improve survival outcomes, as melanoma can often be treated successfully when diagnosed at an early stage. Traditionally, dermatologists rely on dermoscopic visual inspection to identify suspicious lesions by examining asymmetry, border irregularities, color distribution, and textural patterns. However, dermoscopy presents several challenges: diagnostic accuracy varies significantly based on clinical expertise, visual interpretation is subjective, and early melanoma often mimics benign nevus patterns, increasing the likelihood of misdiagnosis. Inter-observer variability and inconsistent evaluations across practitioners further limit the reliability of conventional screening. Additionally, the scarcity of dermatology specialists in many regions and the rising patient load highlight the need for scalable and objective diagnostic solutions that can support timely and accurate decision-making.

Deep learning (DL) has emerged as a transformative approach for dermoscopic image analysis, offering unprecedented capabilities in automated feature extraction, pattern recognition, and clinical decision support. Convolutional neural networks (CNNs) have shown remarkable success in learning hierarchical representations of skin lesions [2–5], while Vision Transformers (ViTs) [6] and hybrid CNN-Transformer architectures demonstrate strong performance in capturing both local texture and global structural information. Recent advancements such as self-supervised learning, multimodal frameworks integrating metadata or clinical notes, and large-scale foundation models, have further accelerated progress in achieving dermatologist-level accuracy [7],[8]. With their potential to enable early disease detection, reduce diagnostic bias, and support large-scale screening programs, DL-based systems are rapidly shaping the future of dermatological diagnostics. This review provides a comprehensive analysis of existing methods, key challenges, dataset limitations, evaluation practices, and promising research directions toward building clinically robust, explainable, and widely deployable AI systems for skin cancer classification [9–11], [12].

The remainder of this review is structured as follows. Section 2 outlines the fundamental background of skin cancer, its clinical characteristics, and commonly used dermoscopic datasets. Section 3 discusses key preprocessing, augmentation, and segmentation techniques essential for improving model reliability. Section 4 provides an in-depth analysis of deep learning approaches, including CNN-based architectures, Vision Transformers, and hybrid models. Section 5 presents open challenges such as data imbalance, domain shift, explainability, and clinical deployment barriers. Finally, Section 6 concludes the review with promising future research directions aimed at developing robust, transparent, and clinically applicable AI systems for skin cancer classification.

A. Background of Skin Cancer, Clinical Characteristics, and Dermoscopic Datasets

Skin cancer is one of the most frequently diagnosed cancers globally and primarily arises due to abnormal proliferation of skin cells, often triggered by prolonged ultraviolet (UV) radiation exposure [13]. It broadly includes melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), each differing in severity, growth patterns, and treatment strategies. Among these, melanoma is the most lethal due to its high metastatic potential, while BCC and SCC, though more common, typically exhibit slower progression. Early detection plays a critical role in preventing disease advancement, making accurate and reliable diagnostic tools essential for improving patient survival outcomes. Clinically, skin cancer diagnosis relies heavily on identifying characteristic visual patterns such as asymmetry, irregular borders, heterogeneous color distribution, surface texture, and lesion elevation. Dermoscopy enhances these observations by revealing subsurface structures that are not visible to the naked eye, making it an indispensable tool for dermatologists. The ABCD rule (Asymmetry, Border, Color, and Diameter), the seven-point checklist, and pattern analysis provide structured frameworks for interpreting dermoscopic features. These characteristics often serve as the foundation for developing machine learning and deep learning models, enabling automated systems to mimic expert visual reasoning and distinguish malignant from benign lesions.

B. Skin Cancer datasets

1) ISIC Archive: The International Skin Imaging Collaboration (ISIC) Archive is the largest publicly available repository for dermoscopic images and serves as the primary benchmark for skin cancer classification research. It contains more than 70,000 dermoscopic images collected from multiple international clinical centers, ensuring diversity in demographic characteristics, acquisition devices, and lesion types. The dataset includes detailed annotations such as lesion diagnosis, segmentation masks, metadata, and clinical notes, making it suitable for tasks ranging from lesion segmentation to malignancy classification. By providing standardized, expert-reviewed ground truth labels, ISIC enables fair comparisons across different machine learning models. Its large scale helps deep learning systems generalize better, though challenges such as class imbalance and varying image quality persist.

2) HAM10000 Dataset: The HAM10000 (Human against Machine with 10,000 images) dataset is one of the most widely used subsets within the ISIC archive. It contains 10,015 dermoscopic images representing seven clinically relevant classes: melanoma, melanocytic nevus, benign keratosis, BCC, dermatofibroma, vascular lesions, and intraepithelial carcinoma. The dataset integrates images from different sources and populations, including professional dermoscopic systems and standard imaging workflows. This diversity enhances its representativeness of real-world clinical conditions. HAM10000's high-resolution images and balanced representation of common lesion types make it especially valuable for training CNNs, Vision Transformers, and hybrid architectures. However, the dataset still exhibits underrepresentation of minority classes such as melanoma and dermatofibroma, requiring rebalancing techniques or data augmentation.

3) ISIC 2017 Challenge Dataset: The ISIC 2017 dataset was introduced as part of the ISIC Challenge for the tasks of lesion segmentation, dermoscopic feature detection, and lesion classification. It includes 2,750 dermoscopic images, providing a balanced platform for evaluating early deep learning methods. The challenge significantly advanced research on segmentation networks like U-Net and early CNN-based classifiers for melanoma detection. Although relatively small compared to newer datasets, ISIC 2017 remains important historically and methodologically due to its precise annotations and task-specific structure. Many early state-of-the-art models were benchmarked here, helping shape the development of automated dermatology systems.

4) ISIC 2018 Challenge Dataset: The ISIC 2018 dataset expanded the scope and scale of the challenge tasks by offering 10,015 images for lesion classification and 2,597 images for segmentation tasks. It introduced more comprehensive task definitions including multi-class classification and lesion attribute detection. The dataset includes high-quality, expert-validated labels and segmentation masks, enabling multimodal research where classification benefits from structural lesion boundaries. ISIC 2018

also fueled the development of attention-based models, multi-branch CNNs, and domain adaptation techniques due to its diverse lesion types and varying acquisition conditions.

5) ISIC 2019 Challenge Dataset: The ISIC 2019 dataset is one of the largest lesion classification benchmarks, containing 25,331 dermoscopic images categorized into eight diagnostic classes. Its increased scale and more heterogeneous data sources simulate real-world clinical variability. With a high imbalance across classes especially melanoma and SCC—the dataset challenges researchers to design robust class-balancing techniques, cost-sensitive losses, and augmentation strategies. ISIC 2019 is frequently used for training large-scale deep learning models, foundation models, and self-supervised learning approaches because of its size and diversity.
6) PH2 Dataset: The PH2 dataset is a smaller but highly curated dataset consisting of **200 dermoscopic images**, focusing on melanoma, atypical nevi, and common nevi. Each image is high resolution and accompanied by expert-provided segmentation masks, clinical borders, and dermoscopic attributes. PH2 is often used for evaluating fine-grained lesion classification, boundary detection, and segmentation-based classification pipelines. Its controlled imaging conditions make it ideal for studying models that require clean and artifact-free inputs, although the limited size restricts its usefulness for deep learning without transfer learning.
7) Derm7pt Dataset: The Derm7pt dataset is designed specifically for clinically interpretable machine learning approaches. It includes 1,011 images annotated using the seven-point checklist, a standard dermatological diagnostic framework. Alongside dermoscopic images, it provides clinical metadata such as diameter, elevation, patient history, and morphological features. This makes the dataset valuable for multimodal learning, explainability research, and building models that emulate clinical diagnostic reasoning. While its size is modest, its structured annotations enable training models that align more closely with dermatological principles.
8) MED-NODE Dataset: The MED-NODE dataset contains 170 images categorized into melanoma and nevus cases. Although small, it is frequently used for classical machine learning, feature-engineering methods, and baseline comparisons for lightweight CNNs. The dataset's manually annotated labels and straightforward binary classification task make it useful for quick experimentation, but its limited size and lack of diverse lesion types restrict its suitability for large-scale deep learning without external pre training.

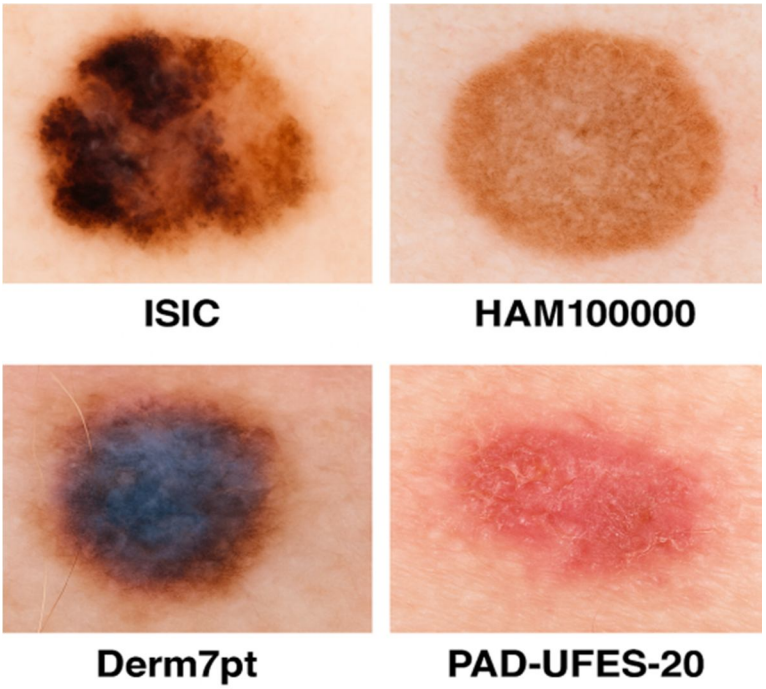


Fig. 1 Preprocessing, Augmentation, and Segmentation Techniques

Deep learning performance in skin cancer classification heavily depends on the quality, consistency, and relevance of input dermoscopic images. This section presents the essential preprocessing steps, augmentation strategies, and segmentation approaches that collectively enhance model reliability, robustness, and generalization across diverse clinical environments.

II. PREPROCESSING TECHNIQUES

Preprocessing is a critical step aimed at standardizing dermoscopic images and reducing noise introduced during acquisition. Common practices include color constancy correction (e.g., Shades-of-Gray, Gray-World) to minimize illumination variations, contrast enhancement to highlight lesion structures, and artifact removal to eliminate distracting elements such as hair, ruler markers, ink spots, and shadows. Methods like DullRazor or morphological filtering are frequently used for hair removal, while median or bilateral filtering helps suppress sensor noise [14]. Additionally, image resizing, normalization, and intensity scaling ensure consistent input formats for CNNs, Vision Transformers, and hybrid architectures. These preprocessing steps establish a clean and uniform image representation, enabling more stable feature extraction and improved downstream classification performance.

A. Data Augmentation

Data augmentation is essential for addressing class imbalance, overfitting, and limited dataset diversity common challenges in skin cancer classification. Traditional augmentation techniques include geometric operations such as rotation, flipping, zooming, cropping, and translation, which expose the model to a wider variety of lesion orientations and shapes. Color-based augmentations, including brightness adjustment, hue shifts, and contrast jittering, improve robustness to lighting and device variations. Advanced augmentation methods like CutMix, MixUp, and Random Erasing introduce regularization effects that enhance generalization. Recently, generative augmentation using GANs [15] or diffusion models has gained attention for synthesizing realistic lesion images, especially for underrepresented melanoma classes. These augmentation strategies collectively help models generalize better across real-world imaging conditions.

B. Lesion Segmentation

Segmentation plays a vital role in isolating the lesion from surrounding healthy skin, ensuring that feature extraction focuses on medically relevant regions. Classical techniques such as thresholding, region growing, and active contours provide basic lesion boundary detection but may struggle with complex textures or low contrast. Deep learning–based approaches, particularly U-Net, ResUNet, and transformer-based segmentation models, achieve significantly higher accuracy by learning contextual and structural patterns directly from annotated data. Segmentation masks can be used to crop the lesion, generate attention maps, or refine classification inputs, thereby reducing background interference and improving interpretability. By accurately delineating lesion borders, segmentation enhances both the precision of feature learning and the clinical reliability of the overall diagnostic pipeline.

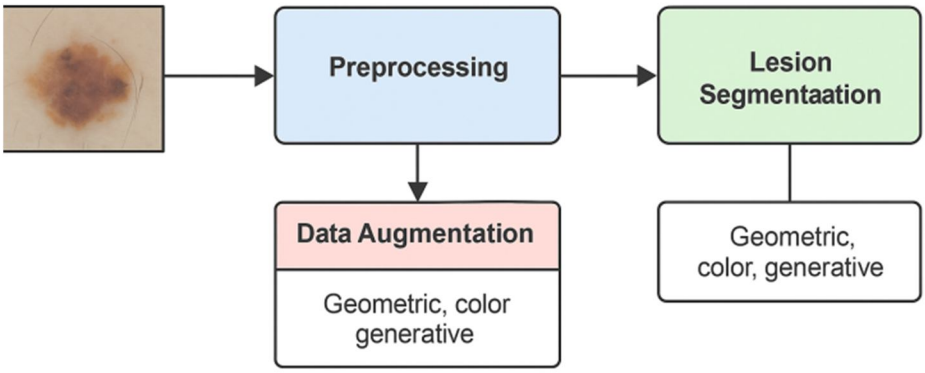


Fig.2 Preprocessing, Augmentation and Segmentation

III. SKIN CANCER CLASSIFICATION USING CNN AND VISION TRANSFORMERS

Convolutional Neural Networks (CNNs) have long been the backbone of skin cancer classification due to their strong ability to extract fine-grained local features from dermoscopic images. Their hierarchical structure enables the network to learn low-level details such as color gradients, lesion edges, and texture patterns in early layers, progressively capturing more complex visual attributes like lesion shape, asymmetry, and irregular borders in deeper layers. This localized feature extraction is particularly valuable for skin lesion analysis, where subtle texture variations and pigment distributions play a crucial role in distinguishing malignant from benign categories. CNN-based models such as ResNet, DenseNet, and MobileNet have demonstrated strong performance, especially when combined with augmentation and transfer learning to deal with limited labeled dermatology datasets.

Vision Transformers (ViTs) and transformer-based vision models introduce a powerful complementary capability by modeling long-range global dependencies across the entire lesion region. Unlike CNNs, which prioritize local receptive fields, ViTs divide the dermoscopic image into patches and use self-attention mechanisms to understand spatial relationships between distant regions. This allows the model to capture high-level semantic context such as lesion symmetry, border irregularity spread, structural heterogeneity, and global color distribution. Such global reasoning is particularly important in skin cancer diagnosis, where the malignancy often depends on holistic patterns rather than isolated textures. Recent transformer variants—such as Swin Transformer, DeiT, and hybrid CNN-ViT architectures—have shown significant improvements in robustness, generalization, and interpretability, making them promising candidates for reliable, scalable skin cancer classification systems.

Table 1. Skin Cancer Classification using CNN and Vision Transformers

Aspect	CNN (Convolutional Neural Networks)	Vision Transformers (ViT / Vision Models)
Feature Extraction	Learns local textures (edges, colors, borders)	Learns global semantic relationships via self-attention
Receptive Field	Grows gradually with depth	Global receptive field from the first layer
Texture Sensitivity	Very good at detecting pigment networks, streaks, dots	Models overall lesion structure better
Handling Complex Shapes	Limited unless deep	Naturally models complex shapes and spatial patterns
Data Requirement	Works well with small datasets (transfer learning)	Needs more data unless optimized (DeiT, Swin)
Robustness to Noise	Sensitive to hair/artifacts	More robust due to long-range global reasoning
Interpretability	Grad-CAM highlights local areas	Attention maps show lesion-level relevance
Computational Cost	Lower; lightweight models available	Higher due to multi-head attention
Parallelization	Efficient GPU/TPU convolution operations	Attention is heavier but improving
Training Stability	Very stable and mature ecosystem	Needs careful optimization (warmup, large batch)
Generalization	Good but may overfit small datasets	Strong cross-domain generalization
Handling Long-Range Dependencies	Weak; dominated by local bias	Strong; connects distant regions directly
Scalability	Performance saturates at extreme depth	Scales extremely well with model size
Clinical Relevance	Captures micro-patterns crucial for melanoma	Captures global asymmetry and border structure

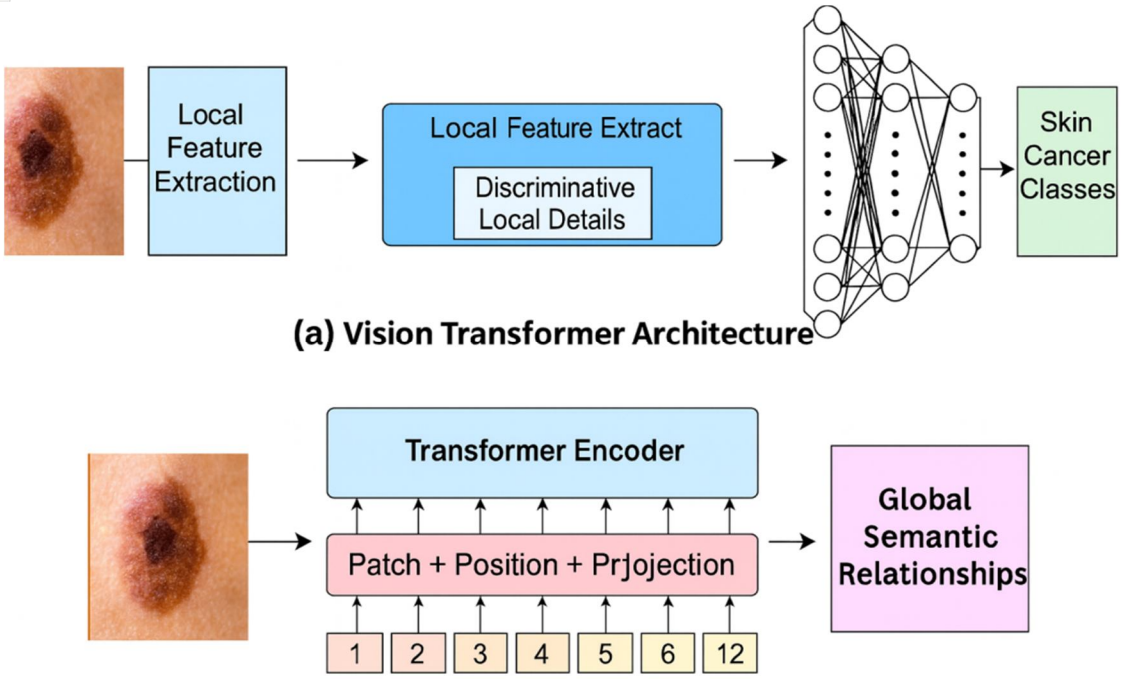


Fig. 2 Vision Transformer architecture

Table 2. Related works to skin cancer transformers

Author	Model	Methodology	Datasets	Performance Metrics
Ding et al. [16]	Deep Attention Branch Network (DABN)	Entropy-based Loss Weighting addresses class imbalance in skin lesion datasets by modifying loss weights. In DABN, Attention branches generate Class Activation Maps (CAMs) during training, emphasizing discriminative regions of lesions.	ISIC-2016 dataset	Average AUC (0.836)
			ISIC-2017 dataset	Average AUC (0.922)
Wei et al. [17]	Densenet-161 + Dual Attention Mechanism	The proposed model integrates two auxiliary supervision branches and KL regularization to enhance the network's capabilities.	ISIC 2017	Accuracy (87.5%) AUC (0.886) Sensitivity (70.9%)
			ISIC 2018	Accuracy (89.00%) AUC (0.976) Sensitivity (83.53%)
			ISIC 2019	Accuracy (89.6%) AUC (0.983) Sensitivity (81.73%)

Qian et al. [18]	CNN + Grouping of Multi-Scale Attention Blocks	The Grouping of Multi-Scale Attention Blocks (GMAB) is designed to extract multi-scale fine-grained features Issan adaptive loss-weighted cross-entropy loss functionassigns different penalties for misclassifications based on a cost-sensitive matrix.	HAM10000 dataset	Accuracy (91.6%) Specificity (96.4%) Sensitivity (73.5%) AUC (0.971)
Hu et al. [2]	Modified EfficientNetV2 + multi-scale fusion structure	The multi-scale fusion structure allows for the extraction of features at multiple levels, from low-level fine-grained details to high-level abstract patterns, through a multi-branch architecture. It can assist network to understand and classify image features, considering both local and global patterns.	HAM10000 dataset	Accuracy (94.3%), Specificity (91.2%), Precision (91.2%), F1 Score (91.3%), AUC (0.993)
			ISIC2019	Accuracy (89.8%), Sensitivity (82.2%), Specificity (98.00%), Precision (88.1%), F1 Score (86.5%), AUC (0.953)
Naveed et al. [19]	DenseNet-121 + Progressive class-wise attention	A class-wise attention mechanism is proposed which processes input tensors through convolution, batch normalization, and ReLU activation to distinguish features among classes. Scores for each class are computed, and a class-wise attention map highlights important regions for classification. Finally, focal loss is used to address class imbalance during training.	HAM10000 dataset	Accuracy (95.8%), AUC (0.997), F1 Score (95.7%)
			ISIC2019	Accuracy (94.9%), AUC (0.994), F1 Score (94.7%)
Li et al. [20]	CNN + Residual Cosine Similarity Attention + Transformer branch	The BC-FCU module integrates local features and global representations from two branches using bidirectional convolution. The RCSA module analyzes semantic information between	ISIC 2018	Accuracy (87.39%)
			XJUSL (privateclinical skin lesion dataset)	Accuracy (84.27%)

		convolutional feature maps and residual edge feature maps using cosine similarity.		
Wan et al. [21]	Long Attention Networks	A long attention network is developed that utilizes the output of the last layer as an attention mask for shallower layers to improve contextual knowledge of the image.	ISIC 2017	AUC (0.937)
			SIIM-ISIC 2020	AUC (0.926)
Wang et al. [22]	Pre-Trained Model + Discriminative feature attention network	The distinguishing trait Attention network generates attention-enhanced features from original features, which are integrated for classification. Misclassification costs vectors direct the attention network to adjust focus based on class relevance, giving minority classes greater attention.	Cifar100	Accuracy (76.77%) Precision (80.42%) Sensitivity (76.77%) F1 Score (76.19%)
			Tiny200	Accuracy (70.57%) Precision (75.47%) Sensitivity (75.57%) F1 Score (70.15%)

IV. DISCUSSION

The comparative diagram highlights the fundamental differences in how Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) analyze dermoscopic images for skin cancer classification. CNNs primarily learn local texture-level information, which makes them highly effective for identifying fine-grained patterns such as pigment networks, edges, streaks, and color irregularities that often characterize early malignant changes. This local inductive bias allows CNNs to perform well even with limited data, which is common in medical imaging. Their hierarchical structure gradually expands the receptive field, but the feature representation still remains largely focused on neighborhood-level interactions. As a result, CNNs excel in capturing micro-level lesion details but may struggle to fully understand global structural relationships, such as asymmetry, shape irregularity, and distributed color gradients that may span the entire lesion area.

In contrast, Vision Transformers operate on a fundamentally different principle by dividing the image into patches and modeling their relationships using self-attention mechanisms. This architecture enables ViTs to capture long-range global dependencies from the very first layer, making them well-suited for analyzing holistic lesion patterns. Features such as border irregularity, global asymmetry, multi-region color variation, and structural layout are more effectively represented in transformer-based models. This global reasoning ability is particularly important in melanoma detection, where clinical diagnosis often depends on assessing the lesion as a whole. However, ViTs generally require larger datasets to learn these representations effectively. When sufficient training samples and appropriate regularization strategies are available, they often outperform CNNs in terms of generalization and robustness, achieving a more comprehensive understanding of lesion morphology.

The discussion also emphasizes that both architectures possess complementary strengths. CNNs offer precise local feature extraction, while ViTs provide powerful global context understanding. For medical imaging tasks like skin cancer classification—where both local details (e.g., atypical dots, texture distortion) and global cues (e.g., asymmetry, border distribution) are equally significant—hybrid models that integrate CNN and ViT modules can leverage the advantages of both paradigms.

Such a fusion strategy has strong potential to enhance sensitivity, reduce misclassification, and improve the model's reliability across diverse lesion types and imaging conditions. Thus, the comparison suggests that future advancements in skin cancer classification may rely on multi-scale, hybrid CNN-Transformer frameworks that balance efficiency, interpretability, and clinical relevance.

A. *Challenges*

1) Class Imbalance: Malignant lesions such as melanoma are significantly underrepresented compared to benign nevi in most datasets. This imbalance causes models to be biased toward majority classes, leading to high accuracy but poor sensitivity for melanoma detection.
2) Lack of Diversity in Skin Tones: Most dermoscopic datasets are dominated by lighter Fitzpatrick skin types, resulting in lower generalizability for darker skin tones. This lack of representation can cause diagnostic disparities and reduce fairness in AI-driven clinical systems.
3) Domain Shift across Imaging Devices: Variations in camera types, dermoscopic equipment, illumination, magnification, and clinical workflows introduce distribution shifts. Models trained on one dataset often perform poorly when evaluated on images from different hospitals or devices.
4) Explainability and Clinical Trust: Clinical adoption requires transparent and interpretable predictions. Dermatologists must understand why a model assigns a certain label. Existing explainability methods (e.g., Grad-CAM, LIME) are not always consistent or clinically meaningful.
5) Need for Lightweight, Mobile-Compatible Models: Deploying AI systems in real-world settings requires efficient models that run on smartphones or low-power devices. Many state-of-the-art networks are computationally heavy and unsuitable for point-of-care or remote clinical environments.
6) Ethical, Fairness, and Bias-Related Issues: AI systems must ensure equitable performance across demographic groups. Bias in training data, privacy concerns, and lack of transparency can lead to ethical risks, misdiagnosis, and reduced trust among clinicians and patients.

B. *Future Research Directions*

1) Development of Lightweight Mobile AI Systems: Designing efficient CNN, transformer, and hybrid architectures optimized for mobile and edge devices will support real-time screening in clinics and underserved regions without requiring high computational resources.
2) Federated Learning for Privacy-Preserving Diagnosis: Federated learning allows hospitals to collaboratively train models without sharing raw patient data. This protects patient privacy and enables large-scale learning from diverse global datasets, improving model robustness.
3) Multimodal Fusion (Image + Clinical Text + Metadata): Combining dermoscopic images with metadata such as age, lesion location, patient history, and textual clinical descriptions can improve predictive performance and align the model more closely with dermatologist decision-making.
4) Diffusion Models for Data Augmentation: Generative diffusion models can synthesize realistic rare-class images (e.g., melanoma or uncommon lesions), helping combat class imbalance, enrich training data, and improve model generalization under limited annotated samples.
5) Improved Explainability Techniques: Advanced explainability tools such as Grad-CAM++, SHAP, integrated gradients, and counterfactual explanations can provide clearer, more trustworthy visual and textual rationales for predictions, increasing clinician acceptance.
6) Dermatology Foundation Models: Large-scale foundation models trained on millions of dermoscopic and clinical images can serve as universal backbones for segmentation, classification, detection, and report generation. These models promise stronger generalization, reduced annotation effort, and broader clinical utility.

V. CONCLUSION

Skin cancer remains a major global health challenge, and early detection is crucial for improving survival outcomes, particularly for aggressive forms such as melanoma. Deep learning has transformed the landscape of dermoscopic image analysis by enabling automated, accurate, and scalable diagnostic systems that increasingly approach dermatologist-level performance.

Through this review, we summarized key skin cancer types, widely used datasets, preprocessing pipelines, and state-of-the-art DL architectures ranging from CNNs and Vision Transformers to hybrid, multimodal, and self-supervised frameworks. While significant progress has been made, real-world deployment is still hindered by issues such as class imbalance, limited skin tone diversity, domain shift, explainability gaps, and ethical concerns. Future advancements such as lightweight mobile AI, federated learning, multimodal fusion, diffusion-based augmentation, improved interpretability, and large dermatology foundation models hold considerable promise for bridging the gap between research and clinical practice. Ultimately, the development of robust, transparent, and equitable AI systems will be essential to support dermatologists, enhance early diagnosis, and enable widespread, accessible skin cancer screening across diverse populations.

REFERENCES

[1] Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., Henkel, J.: A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. IEEE Access. 12, 41180–41218 (2024). https://doi.org/10.1109/ACCESS.2024.3376441.

[2] Hu, Z., Mei, W., Chen, H., Hou, W.: Multi-scale feature fusion and class weight loss for skin lesion classification. Comput. Biol. Med. 176, (2024). https://doi.org/10.1016/j.compbiomed.2024.108594.

[3] Girepunje, S., Singh, P.: ECAM-Net: A Lightweight Convolutional Neural Network Enhanced with Efficient Channel Attention for Lung Cancer Classification. Procedia Comput. Sci. 260, 126–133 (2025). https://doi.org/10.1016/j.procs.2025.03.185.

[4] Girepunje, S., Singh, P.: A Transfer Learning-Based DenseNet with Multi kernel aggregated Channel Attention Block for COVID-19 Classification using CT scan images. 2024 OPJU Int. Technol. Conf. Smart Comput. Innov. Adv. Ind. 4.0, OTCON 2024. 1–6 (2024). https://doi.org/10.1109/OTCON60325.2024.10687731.

[5] Annappa Wei-Chiang Hong Dasari Haritha Lavanya Devi Editors, B.G.: Lecture Notes in Electrical Engineering 1262 High Performance Computing, Smart Devices and Networks Select Proceedings of CHSN 2023. (2023).

[6] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., Tao, D.: A Survey on Vision Transformer. IEEE Trans. Pattern Anal. Mach. Intell. 45, 87–110 (2023). https://doi.org/10.1109/TPAMI.2022.3152247.

[7] de Santana Correia, A., Colombini, E.L.: Attention, please! A survey of neural attention models in deep learning. Springer Netherlands (2022). https://doi.org/10.1007/s10462-022-10148-x.

[8] Hassanin, M., Anwar, S., Radwan, I., Khan, F.S., Mian, A.: Visual Attention Methods in Deep Learning: An In-Depth Survey. 1–20 (2022).

[9] Park, J., Woo, S., Lee, J.Y., Kweon, I.S.: BAM: Bottleneck attention module. Br. Mach. Vis. Conf. 2018, BMVC 2018. (2019).

[10] Liu, Y., Shao, Z., Teng, Y., Hoffmann, N.: NAM: Normalization-based Attention Module. (2021).

[11] Qiao, Z., Yuan, X., Zhuang, C., Meyarian, A.: Attention pyramid module for scene recognition. Proc. - Int. Conf. Pattern Recognit. 7521–7528 (2020). https://doi.org/10.1109/ICPR48806.2021.9412235.

[12] Adil, N., Singh, P., Nagwani, N.K.: Interpretable Lightweight CNN for Colon and Lung Cancer Classification with LIME Based Explainability. Proc. - 2024 IEEE Int. Conf. Intell. Syst. Smart Green Technol. ICISSGT 2024. 122–127 (2024). https://doi.org/10.1109/ICISSGT58904.2024.00034.

[13] Tiwari, A., Ghose, A., Hasanova, M., Faria, S.S., Mohapatra, S., Adeleke, S., Boussios, S.: The current landscape of artificial intelligence in computational histopathology for cancer diagnosis. Discov. Oncol. 16, (2025). https://doi.org/10.1007/s12672-025-02212-z.

[14] Jain, A., Singh, P.: Optimized Lightweight Network for Enhanced Image Denoising. 2025 IEEE Int. Conf. Interdiscip. Approaches Technol. Manag. Soc. Innov. IATMSI 2025. 3, 1–6 (2025). https://doi.org/10.1109/IATMSI64286.2025.10984546.

[15] Kumar, A., Singh, D.: Generative Adversarial Network-Based Augmentation with Noval 2-step Authentication for Anti-coronavirus Peptide Prediction. IEEE/ACM Trans. Comput. Biol. Bioinforma. 21, 1942–1954 (2024). https://doi.org/10.1109/TCBB.2024.3431688.

[16] Ding, Y., Ma, Z., Wen, S., Xie, J., Chang, D., Si, Z., Wu, M., Ling, H.: AP-CNN: Weakly Supervised Attention Pyramid Convolutional Neural Network for Fine-Grained Visual Classification. IEEE Trans. Image Process. 30, 2826–2836 (2021). https://doi.org/10.1109/TIP.2021.3055617.

[17] Wei, Z., Li, Q., Song, H.: Dual attention based network for skin lesion classification with auxiliary learning. Biomed. Signal Process. Control. 74, 103549 (2022). https://doi.org/10.1016/j.bspc.2022.103549.

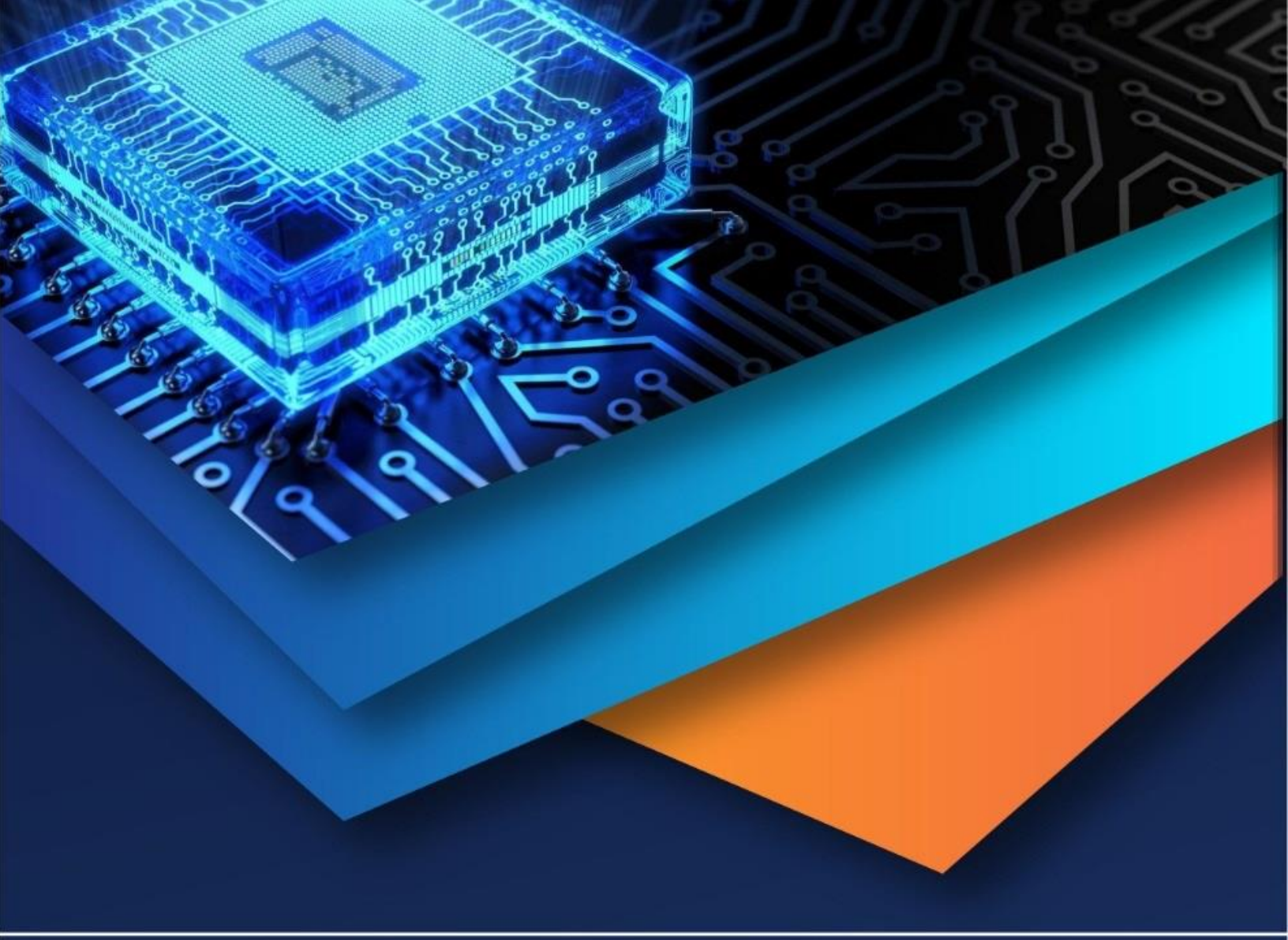
[18] Qian, S., Ren, K., Zhang, W., Ning, H.: Skin lesion classification using CNNs with grouping of multi-scale attention and class-specific loss weighting. Comput. Methods Programs Biomed. 226, 107166 (2022). https://doi.org/10.1016/j.cmpb.2022.107166

[19] Naveed, A., Naqvi, S.S., Khan, T.M., Razzak, I.: PCA: Progressive class-wise attention for skin lesions diagnosis. Eng. Appl. Artif. Intell. 127, 107417 (2024). https://doi.org/10.1016/j.engappai.2023.107417.

[20] Li, A., Zhang, D., Yu, L., Kang, X., Tian, S., Wu, W., You, H., Huo, X.: Residual cosine similar attention and bidirectional convolution in dual-branch network for skin lesion image classification. Eng. Appl. Artif. Intell. 133, 108386 (2024). https://doi.org/10.1016/j.engappai.2024.108386.

[21] Wan, Y., Cheng, Y., Shao, M.: MSLANet: multi-scale long attention network for skin lesion classification. Appl. Intell. 53, 12580–12598 (2023). https://doi.org/10.1007/s10489-022-03320-x.

[22] Wang, L., Zhang, L., Qi, X., Yi, Z.: Deep Attention-Based Imbalanced Image Classification. IEEE Trans. Neural Networks Learn. Syst. 33, 3320–3330 (2022). https://doi.org/10.1109/TNNLS.2021.3051721.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)