International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

Automated Classification of Skin Lesions Using Convolutional Neural Networks

Dr. Sharad Mathur[1], Dr Ashish Rai[2]

[1]Associate Professor, Department of Computer Science, Lachoo Memorial College Sc & Tech, Jodhpur
[2]Assistant Professor, Department of Computer Science, Lachoo Memorial College Sc & Tech, Jodhpur

Abstract

The incidence of skin cancer is a growing global health concern, making early and accurate detection essential for patient survival. While traditional diagnosis relies heavily on the visual inspection of dermoscopic images by expert dermatologists, this process is highly subjective and time-consuming. Convolutional Neural Networks (CNNs) have emerged as powerful tools for automating this diagnostic workflow. This paper presents a comprehensive review and architectural analysis of automated skin lesion classification using CNNs. We detail the end-to-end pipeline, including advanced preprocessing techniques like the DullRazor algorithm for artifact removal, and evaluate the performance of state-of-the-art networks such as ResNet, EfficientNet, and hybrid Vision Transformers on benchmark datasets like HAM10000. Furthermore, we address the critical barriers to clinical deployment: the "black-box" nature of neural networks, mitigated through Explainable AI (XAI) frameworks like Grad-CAM and SHAP, and the persistent demographic biases tied to skin tone representation. By synthesizing these elements, this paper provides a roadmap for developing transparent, equitable, and highly accurate dermatological clinical decision support systems.

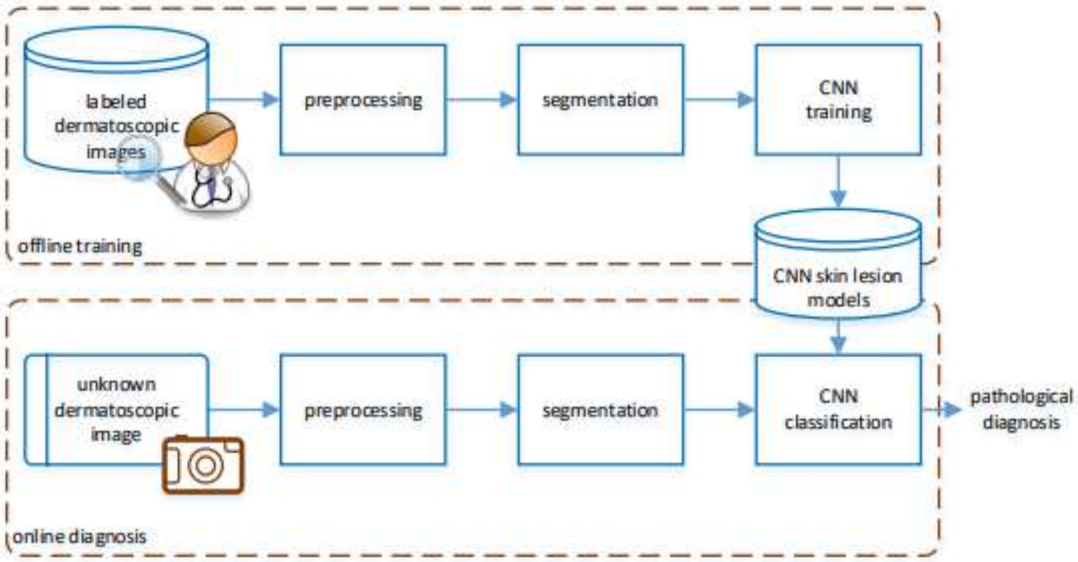
Keywords: Convolutional Neural Networks, Skin Lesion Classification, Dermoscopy, Explainable AI, Algorithmic Fairness, Melanoma.

1. Introduction

Skin cancer remains one of the most widely diagnosed malignancies globally, with millions of new cases reported annually. The pathologies range from common, non-invasive lesions like basal cell carcinoma (BCC) to malignant melanoma, which accounts for the vast majority of skin cancer-related fatalities.[1] The prognosis for melanoma is heavily dependent on the timeline of detection; if excised in its early stages, the five-year survival rate exceeds 95%.[2]

Historically, dermatologists have relied on dermoscopy—a non-invasive imaging technique that magnifies subsurface skin structures—to identify malignancies. However, human interpretation of these complex morphological patterns (such as atypical pigment networks and irregular borders) is subjective and prone to inter-observer variability. To address this bottleneck, Artificial Intelligence (AI) and Machine Learning (ML) have been increasingly integrated into the diagnostic process.[3]

Deep learning, specifically through Convolutional Neural Networks (CNNs), allows computational systems to automatically learn hierarchical feature representations directly from raw pixel data, bypassing manual feature engineering. While early CNNs demonstrated the ability to match board-certified dermatologists in constrained laboratory settings, transitioning these models into robust clinical tools requires overcoming several hurdles: severe class imbalances in training data, the presence of physical artifacts (hair, surgical ink), opaque decision-making processes, and significant performance disparities across different human skin tones.[5]



This paper explores the current state-of-the-art in automated skin lesion classification. Section II examines foundational dataset ecosystems. Section III details the preprocessing and architectural methodologies. Section IV presents experimental benchmarks. Section V discusses Explainable AI (XAI), and Section VI tackles algorithmic fairness.[6]

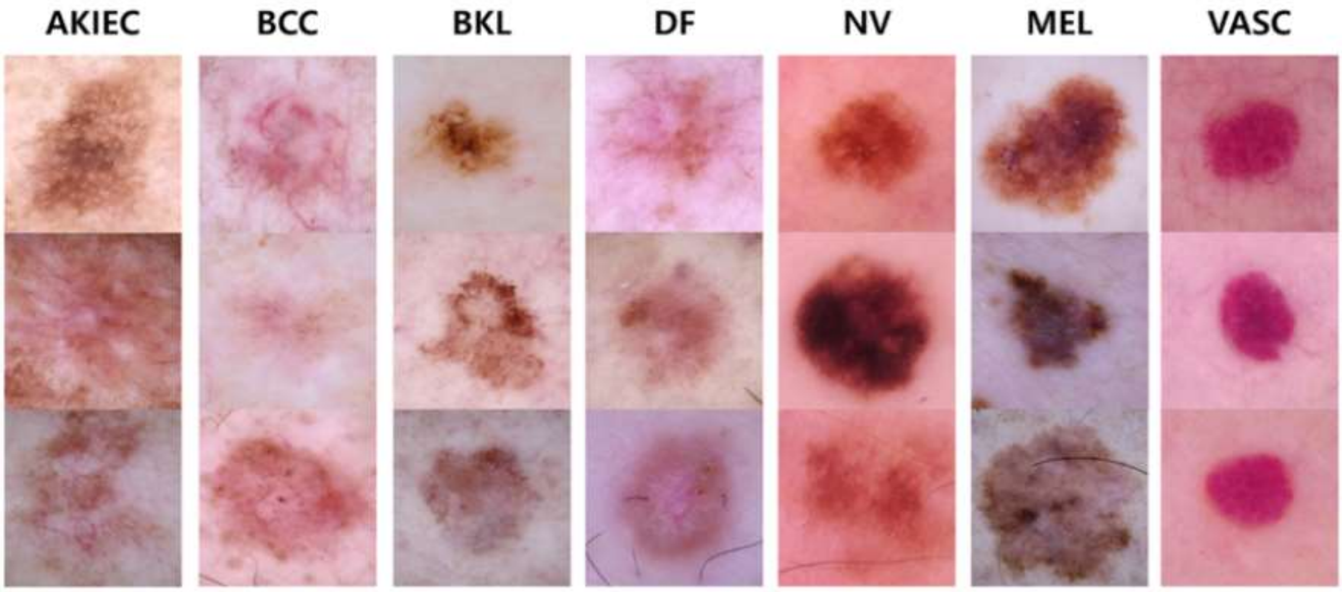
2. Dataset Ecosystems and Clinical Complexity

The performance of any deep learning model is intrinsically tied to the quality, volume, and diversity of its training data. In computational dermatology, a few key repositories drive the majority of algorithmic development.

2.1 The HAM10000 Benchmark

The HAM10000 (Human Against Machine with 10,000 training images) dataset is the standard benchmark for multi-class skin lesion classification. It consists of 10,015 high-quality dermoscopic images spanning seven diagnostic categories:

- Actinic Keratoses (AKIEC): Pre-malignant lesions caused by sun damage.
- Basal Cell Carcinoma (BCC): A common, locally destructive but rarely metastatic cancer.
- Benign Keratosis-like Lesions (BKL): Benign lesions that often mimic melanoma.
- Dermatofibroma (DF): Benign dermal nodules often resulting from minor trauma.
- Melanoma (MEL): Highly aggressive, malignant tumors originating from melanocytes.
- Melanocytic Nevi (NV): Common benign moles.
- Vascular Lesions (VASC): Benign vascular malformations like cherry angiomas.



A primary challenge with HAM10000 is its severe class imbalance. For example, Melanocytic Nevi (NV) make up roughly 67% of the dataset, while rare classes like Dermatofibroma (DF) account for just over 1%. Unoptimized models trained on this data tend to collapse, over-predicting the majority class while failing to identify critical minority classes like melanoma.[7]

2.2 The ISIC Archives and 3D Total Body Photography

The International Skin Imaging Collaboration (ISIC) hosts annual challenges that push the boundaries of the field. While earlier datasets (2016-2020) focused on dermoscopic images, the ISIC 2024 challenge introduced the SLICE-3D dataset. This dataset shifted focus toward 3D Total Body Photography (TBP), providing over 400,000 image crops that simulate lower-quality, real-world smartphone photos often used in telehealth triage.[8]

3. Methodology and Automated Pipeline

Developing a robust classification model requires a multi-stage pipeline, moving from raw image acquisition to final probabilistic output.

3.1 Artifact Removal: The DullRazor Algorithm

Dermoscopic images are frequently occluded by thick body hair, which introduces sharp artificial gradients that confuse standard convolutional edge-detection filters.[9] The DullRazor algorithm is the industry standard for programmatic hair removal. The steps are as follows:

- Grayscale Conversion: The RGB image is flattened to isolate structural contrasts.
- Morphological Closing: Linear or cross-shaped kernels are applied to bridge the dark, narrow gaps created by hair strands.
- Black-Hat Filtering: The original image is subtracted from the morphologically closed image to isolate the exact coordinates of the hair.
- Mask Generation: Binary thresholding is applied to create a pixel-level hair mask.
- Inpainting: The pixels identified by the mask are excised from the original RGB image and replaced using bilinear interpolation or partial-differential-equation (PDE) based smoothing from surrounding healthy skin.

3.2 Color Constancy and Illumination Normalization

Environmental lighting variations (fluorescent vs. natural daylight) drastically alter the perceived color of a lesion—a critical diagnostic feature. Color constancy algorithms (such as Shades of Gray or Max-RGB) mathematically estimate the environmental illuminant and normalize the pixel matrix, effectively standardizing the image as if it were taken under pure white light.[10]

3.3 Advanced Architectures: From CNNs to Hybrids

The architectural topology of the feature extractor dictates the model's learning capacity:

- **Residual Networks (ResNet)**: By utilizing skip connections, ResNet architectures (like ResNet-50) solve the vanishing gradient problem, allowing for deep networks that capture complex hierarchical features without degrading.
- **EfficientNet**: Models like EfficientNet-B0 utilize a compound scaling coefficient that balances network depth, width, and image resolution simultaneously. They consistently offer state-of-the-art accuracy with significantly fewer computational parameters.
- **Vision Transformers (ViT) and Hybrids**: While CNNs excel at extracting local textures, they struggle with global context (like overall lesion asymmetry). Hybrid models pass CNN-extracted feature maps into Transformer blocks, utilizing self-attention mechanisms to map long-range dependencies across the entire lesion.[11]

4. Experimental Benchmarks

Table presents a comparative performance analysis of prominent deep learning architectures evaluated on the HAM10000 dataset for multi-class classification.

Architecture	Backbone Type	Accuracy	Precision	Recall	F1-Score
MobileNetV2	Depthwise CNN	91.11%	0.79	0.74	0.76
ResNet-50	Residual CNN	93.90%	0.86	0.86	0.86
EfficientNet-B0	Compound CNN	94.87%	0.97	0.93	0.95
ConvNeXt V2	Modern CNN	93.20%	0.81	0.77	0.79
Hybrid (CNN+ViT)	Hybrid	96.41%	0.96	0.95	0.95

Hybrid models consistently demonstrate the highest evaluation metrics by combining the localized texture recognition of CNNs with the global structural awareness of Transformers.[12]

5. Explainable AI (XAI) in Dermatology

A critical barrier to clinical adoption is the "black-box" nature of deep learning. Physicians cannot ethically recommend aggressive biopsies based on an opaque probabilistic score. Explainable AI (XAI) bridges this gap by providing visual and mathematical justifications for the model's outputs.

- **Grad-CAM (Gradient-Weighted Class Activation Mapping)**: Grad-CAM intercepts the feature maps at the final convolutional layer. By calculating the gradient of the target class score, it generates a coarse localization heatmap. If the heatmap aligns with clinical "ABC" criteria

(Asymmetry, Border irregularity, Color), the clinician can trust the output. If it highlights background artifacts (like a ruler), the prediction is discarded.[13]

- **LIME and SHAP**: While Grad-CAM provides spatial localization, frameworks like SHAP (SHapley Additive exPlanations) provide quantitative feature attribution, assigning an exact numerical weight to how much a specific texture or patient metadata variable influenced the final risk score.

6. Algorithmic Fairness and Clinical Equity

Despite high benchmark accuracies, a severe systemic bias exists within the foundational training data regarding human skin tone. Major datasets like ISIC and HAM10000 are predominantly sourced from populations of European descent, resulting in an extreme overrepresentation of lighter skin tones.

6.1 The Measurement Gap

Historically, AI auditing has relied on the Fitzpatrick Skin Type (FST) scale. However, the FST was designed to measure UV burn response, not visual pigmentation, and it severely compresses the variance of darker skin tones.[14] Researchers are now transitioning to objective metrics like the Individual Typology Angle (ITA)—derived from CIELAB color space—and the 10-shade Monk Skin Tone (MST) scale to properly audit model performance.

6.2 Performance Disparities

Empirical audits reveal that models trained on imbalanced data suffer severe performance degradation when applied to darker skin. For instance, diagnostic accuracy often drops significantly for Fitzpatrick types V–VI compared to types I–II. This occurs because CNNs trained on light skin fail to recognize how malignancies visually present on melanin-rich skin (e.g., lower contrast between the lesion and healthy tissue).[15]

Mitigating this bias requires targeted interventions. Synthetic data augmentation using Latent Diffusion Models (LDMs) can generate high-fidelity, synthetic images of malignancies on darker skin, artificially balancing the dataset and forcing the network to learn invariant pathological features.

7. Conclusion

Automated skin lesion classification using Convolutional Neural Networks represents a profound advancement in dermatological oncology. Modern architectures like EfficientNet and Hybrid Vision Transformers, when paired with robust preprocessing pipelines like DullRazor, achieve diagnostic accuracies that rival expert clinicians. However, raw accuracy is insufficient for real-world medical deployment. The integration of Explainable AI (Grad-CAM, SHAP) is mandatory to foster physician trust. Furthermore, the AI research community must urgently address the demographic biases embedded in current datasets by adopting equitable skin-tone measurement scales (ITA, Monk) and leveraging generative augmentation to ensure these life-saving diagnostic tools are effective for all global populations, regardless of skin color.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," CA: A Cancer Journal for Clinicians, vol. 70, no. 1, pp. 7-30, 2020.
2. N. Codella et al., "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," arXiv preprint arXiv:1902.03368, 2019.

3. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, p. 180161, 2018.
4. N. Kurtansky et al., "The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D TBP for skin cancer detection," JEADV, 2024
5. T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "DullRazor: a software approach to hair removal from images," Computers in Biology and Medicine, vol. 27, no. 6, pp. 533-543, 1997.
6. C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1088-1099, 2018.
7. S. Mathur, A. Rai, D. Mathur, " Predictive Data Mining for Disease Diagnosis-Decision Tree Approach", International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181,Vol. 14 Issue 05, May-2025
8. M. Abbas et al., "Enhanced Skin Disease Diagnosis through Convolutional Neural Networks and Data Augmentation Techniques," Journal of Computing & Biomedical Informatics, vol. 7, pp. 87-106, 2024.
9. S. Mathur, A. Rai, "Enhancing Disease Forecasting Through Naïve Bayes Classification Techniques", International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181, Vol. 14 Issue 12 , December - 2025
10. R. Jalaboi, O. Winther, and A. Galimzianova, "Dermatological Diagnosis Explainability Benchmark for Convolutional Neural Networks," arXiv preprint arXiv:2302.12084, 2023
11. S. Mathur, B. Joshi, Application of Naïve Bayes Classification for disease Prediction International Journal of Management, IT & Engineering, Vol. 9 Issue 4, April 2019, ISSN: 2249-0558
12. A. Daneshjou et al., "Disparities in Dermatology AI Performance on a Diverse Clinical Dataset," Nature Medicine, 2022
13. E. Monk, "The Monk Skin Tone Scale," Google AI Research, 2023.
14. S. Mathur, B. Joshi, Survey on Needs, Applications and Agorithms of Data Mining for Healthcare, International Journal of Advance Research in Science and Engineering", Vol 6(8), 80-87, August, 2017, ISSN: 2319-8346.
15. M. Benčević et al., "Understanding skin color bias in deep learning-based skin lesion segmentation," Computer Methods and Programs in Biomedicine, vol. 245, p. 108044, 2024