Automated Multi-Class Skin Lesion Detection Using Deep Learning and Hybrid CNN-Transformer Architectures

Dogga Pavan Sekhar[†]
2K22CSUN01167
Dept. of Computer Science and Engineering
Manav Rachna University
Faridabad, India
[†] doggapavansekhar@gmail.com

Abstract* - *Skin cancer, particularly melanoma, is among the most dangerous forms of cancer due to its rapid progression and high mortality rate if not detected early. Early and accurate detection of skin lesions is therefore crucial for effective treatment and improved patient outcomes. Traditional diagnosis relies heavily on visual inspection by dermatologists, which is time-consuming and subject to inter-observer variability. Recent advancements in deep learning have enabled automated analysis of dermoscopic images, offering the potential to assist clinicians in accurate and efficient diagnosis. This project focuses on developing a robust and accurate skin lesion classification system using deep learning techniques. Leveraging publicly available datasets such as ISIC, the system aims to classify multiple types of skin lesions, including melanoma, with high precision. The proposed approach explores advanced convolutional neural networks (CNNs), hybrid CNN-transformer architectures, and attention-based mechanisms, with the goal of improving classification performance and providing reliable, clinically useful predictions.

I. Problem Statement

Early detection of skin cancer, particularly malignant lesions such as melanoma, is critical for improving patient survival rates. However, accurate diagnosis remains challenging due to subtle visual differences between benign and malignant lesions, variability in lesion size, shape, and color, and the presence of artifacts such as hair and shadows in dermoscopic images. Manual diagnosis by dermatologists is time-consuming and prone to inter-observer variability, leading to potential inconsistencies in clinical assessment.

In India, although skin cancer incidence is relatively lower compared to Western countries, it has been steadily increasing in recent years. According to the National Cancer Registry Programme (ICMR-NCRP 2024), non-melanoma skin cancers account for approximately 0.5–1% of all cancers, while melanoma cases are rising at an estimated 3–5% annually, particularly in states with higher UV exposure. Reports indicate that around 25,000–30,000 new cases of skin cancer are diagnosed each year across the country, with a concerning trend of late-stage detection, which significantly reduces survival outcomes.

While deep learning–based automated detection systems have demonstrated promising results, they often face significant challenges, including class imbalance, limited generalization across diverse datasets, and difficulty in providing interpretable outputs that can be trusted in clinical practice.

This project aims to develop an automated skin lesion detection and classification system capable of accurately identifying and classifying multiple types of skin lesions from dermoscopic images. The proposed system seeks to address key challenges such as data imbalance, feature variability, and clinical interpretability, ensuring high diagnostic accuracy and reliability to assist dermatologists in real-world decision-making and help improve early detection rates in the Indian healthcare context.

II. Introduction

Skin cancer is one of the most common and deadly forms of cancer worldwide, with melanoma being particularly aggressive if not detected early. Early detection and accurate diagnosis are critical to improving patient survival rates. Traditionally, dermatologists diagnose skin lesions by visually examining dermoscopic images or using biopsy tests. However, this process is time-consuming, subjective, and prone to human error, especially in regions with limited access to experienced clinicians. As a result, many patients may not receive timely or accurate diagnosis, leading to delayed treatment and poorer outcomes.

To address this challenge, we propose an automated multi-class skin lesion detection system powered by deep learning. This system aims to classify various types of skin lesions, including melanoma, benign nevi, basal cell carcinoma, and others, using dermoscopic images. By leveraging advanced convolutional neural networks (CNNs), hybrid CNN-transformer architectures, and attention mechanisms, the system can learn both local texture features and global contextual patterns that are critical for distinguishing between lesion types. Additionally, the system is designed to provide interpretable predictions, allowing clinicians to understand the model's focus areas and enhance trust in AI-assisted diagnosis.

However, building an effective automated system for skin lesion detection presents several challenges:

- **High Intra-Class Variability and Low Inter-Class Distinction** – Lesions of the same type can vary significantly in shape, color, and size, while different types of lesions may appear visually similar.
- **Class Imbalance** – Malignant lesions are less frequent than benign lesions, making it difficult for models to learn robust representations without bias.
- **Image Artifacts and Noise** – Dermoscopic images often include hair, shadows, and uneven lighting, which can confuse models if not preprocessed correctly.
- **Generalization Across Datasets** – Models trained on one dataset may not perform well on images from different sources due to variations in imaging conditions and equipment.

To overcome these challenges, our system incorporates:

- **Preprocessing Techniques** such as hair removal, color normalization, and lesion segmentation to improve image quality.
- **Deep Learning Architectures** including EfficientNets, hybrid CNN-Transformer models, and attention mechanisms to capture both local and global features.
- **Advanced Training Strategies** like data augmentation, class reweighting, and mixup to handle class imbalance and improve robustness.
- **Interpretability Methods** such as Grad-CAM and attention visualization to ensure that the model's decisions are clinically meaningful.

By developing this intelligent and robust system, we aim to assist dermatologists in early and accurate diagnosis, reduce the dependency on manual inspection, and improve patient outcomes. The system is designed not only for high accuracy but also for clinical reliability, interpretability, and potential deployment in real-world healthcare settings.

III. Literature Survey

[1] Tschandl et al. (2023), in their study titled "**A Survey on Deep Learning for Skin Lesion Segmentation**", provide a comprehensive overview of the progress and challenges in automating skin lesion segmentation, which is a crucial step in computer-aided diagnosis of skin cancer. The research emphasizes the importance of accurate lesion boundaries for diagnostic criteria such as lesion size, border irregularity, and color variation. Manual segmentation is often tedious and inconsistent, motivating the need for automated approaches. The survey reviews 177 studies conducted between 2014 and 2022, highlighting how deep learning methods have surpassed traditional techniques by jointly learning feature extraction and decision-making from large annotated datasets. It also discusses the evolution of publicly available datasets, including ISIC and HAM10000, annotation techniques, and challenges such as inter- and intra-annotator variability, underlining the importance of reliable ground truths and scalable annotation strategies. The study categorizes deep learning models into encoder–decoder networks (e.g., U-Net), ensemble and hybrid approaches, and emerging transformer-based architectures. Key methodological components are analyzed, including loss functions (Dice, Jaccard, Tversky, adversarial losses), data augmentation strategies, and evaluation metrics (Dice coefficient, Jaccard index, sensitivity, specificity) commonly used to benchmark model performance. Despite significant advances, the paper identifies several persistent challenges: limited availability of annotated data, generalization across diverse imaging domains, model interpretability, and fairness across different skin tones. For future research, the authors suggest exploring semi-supervised or weakly supervised learning, integration of transformer architectures, and improved robustness to image artifacts. Overall, the survey serves as a valuable guide for researchers aiming to develop reliable deep learning models for skin lesion segmentation.

[2] Naqvi et al. (2023), in their study titled "**Skin Cancer Detection Using Deep Learning—A Review**", explore the rapid advancements of deep learning (DL) methods in diagnosing skin cancer and their potential to assist dermatologists in early detection, thereby reducing mortality. The research highlights that traditional diagnostic methods, even with dermoscopy, often struggle with early-stage or atypical melanomas. In contrast, DL-based models—particularly convolutional neural networks (CNNs)—have achieved remarkable accuracy by automatically extracting complex visual patterns from dermoscopic images. The study compares popular architectures such as AlexNet, VGG, ResNet, DenseNet, and MobileNet, emphasizing the benefits of transfer learning, ensemble strategies, and hybrid segmentation–classification approaches for improving performance. Reported accuracies in recent studies range from approximately 76% to over 99%, depending on dataset size, preprocessing techniques (e.g., hair removal, data augmentation, segmentation), and network design. Lightweight models like MobileNet are highlighted for their suitability for real-time and mobile deployment. The review also catalogs key datasets, including HAM10000, ISIC (2016–2020), PH2, Dermofit, BCN20000, and PAD-UFES-20, describing their lesion classes and image diversity. Despite these advances, several challenges remain: limited dataset diversity and class imbalance contribute to bias toward lighter skin tones, while deeper networks require high computational resources, limiting deployment in resource-constrained environments. The authors recommend developing larger, more diverse datasets, energy-efficient architectures, and incorporating explainability and uncertainty quantification to enhance clinical trust. Overall, the paper emphasizes deep learning's transformative potential in automated skin cancer detection while outlining the necessary steps to achieve robust, equitable, and clinically viable solutions.

[3] Patel et al. (2022), in their study titled "**Skin-DeepNet: A Deep Learning Framework for Fully Automated Early Diagnosis and Classification of Skin Cancer**", present a comprehensive deep learning pipeline for early skin cancer

detection, integrating advanced image preprocessing, segmentation, and ensemble classification techniques. The system begins by enhancing dermoscopic images using Adaptive Gamma Correction with Weighting Distribution (AGCWD) along with a specialized hair-removal method, improving lesion visibility while minimizing artifacts. For segmentation, the framework employs a hybrid Mask R-CNN + GrabCut approach, where Mask R-CNN generates pixel-level lesion masks and GrabCut refines boundaries, achieving Intersection over Union (IoU) scores above 99% on both ISIC 2019 and HAM10000 datasets. For classification, Skin-DeepNet uses a dual-feature strategy: a High-Resolution Network (HRNet) backbone combined with an attention mechanism to capture multi-scale lesion features, and a Deep Belief Network (DBN) classifier supported by a Deep Restricted Boltzmann Machine. Outputs from these classifiers are fused through ensemble decision strategies, including boosting (XGBoost) and stacking (Logistic Regression, Random Forest, Extra Trees), to maximize predictive accuracy. Experiments on large multi-class datasets demonstrate exceptional performance, with accuracies of 99.65% on ISIC 2019 and 100% on HAM10000, outperforming existing state-of-the-art methods across metrics such as precision, recall, and AUC. The study highlights Skin-DeepNet's clinical potential for early melanoma detection, while noting future research directions including improving rare class representation, enhancing attention mechanisms, and performing large-scale clinical validations to ensure real-world applicability.

[4] Singh et al. (2022), in their study titled "**Explainable Deep Inherent Learning for Multi-Class Skin Lesion Classification**", propose a novel deep learning framework that combines high diagnostic accuracy with interpretability to enhance trust in automated skin cancer detection. The study addresses the global burden of skin cancer—particularly basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma—noting that traditional diagnostic methods, including visual inspection, the ABCD rule, and dermoscopy, are often limited by image artifacts, variable lighting, and subtle visual similarities between benign and malignant lesions. To overcome these challenges, the authors introduce a deep inherent learning model with 54 layers that employs multiple convolutional filters, inherent blocks, and inner quadrant correlation to strengthen feature extraction and maintain rich information flow, even with scarce or degraded data. The model was trained and tested on the ISIC 2018/HAM10000 dermoscopic dataset, containing over 10,000 images across seven lesion classes. Experimental results demonstrate strong multi-class classification accuracy and robust lesion localization, outperforming shallower networks while reducing error rates. A key contribution of this framework is its integration of explainable AI (XAI) techniques, including saliency maps and guided backpropagation, which visualize the image regions influencing the model's predictions. These visual explanations provide clinicians with transparent evidence for each decision, supporting the model's use as a second-opinion tool in clinical settings. By combining high performance with interpretability, the deep inherent learning framework advances responsible AI in dermatology and provides a reliable, explainable aid for early skin cancer diagnosis.

[5] Sulthana et al. (2024), in their study titled "**A Novel End-to-End Deep Convolutional Neural Network Based Skin Lesion Classification Framework**", propose a comprehensive deep learning framework for accurate classification of skin lesions. The study addresses the challenges posed by skin diseases—including melanoma, nevi, keratosis, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma—which exhibit high visual variability, subtle boundaries, and artifacts such as hair. To tackle these issues, the authors introduce novel image pre-processing techniques, including a modified Gaussian filtering approach for precise lesion segmentation and fractal-based texture analysis (SFTA) for robust feature extraction. For classification, they develop a customized lightweight S-MobileNet CNN, employing depthwise separable convolutions, the Mish activation function for smoother gradient flow, and selective L1-norm pruning to remove insignificant filters, reducing computational cost without compromising accuracy. The framework was evaluated on the HAM10000 dataset, containing 10,000 dermoscopic images across seven lesion types, achieving a high classification accuracy of 98.35%, outperforming standard models such as Inception, ResNet, and MobileNet. This work demonstrates that combining advanced pre-processing with a lightweight, optimized CNN can enable accurate, efficient, and clinically relevant skin lesion classification, supporting dermatologists in early diagnosis and improving patient outcomes.

[6] Nugroho et al. (2024), in their study titled "**Image Dermoscopy Skin Lesion Classification Using Deep Learning Method: Systematic Literature Review**", present a comprehensive review of recent advances in automated skin lesion classification using deep learning techniques. The study underscores the critical need for accurate and early detection of skin cancer, particularly melanoma, due to the visual diversity and subtlety of lesions. Using a systematic literature review approach, the authors analyzed studies from major digital libraries, focusing on datasets, preprocessing, segmentation, and classification methods. Dermoscopic images—primarily from ISIC datasets (ISIC-16 to ISIC-20)—are commonly used because of their high quality and clinical relevance. Key preprocessing techniques include artifact removal (e.g., DullRazor), color enhancement, and contrast adjustment, which improve segmentation and classification performance. Segmentation methods reviewed include thresholding, active contour models, hybrid approaches, and deep learning-based methods such as R-CNN, which effectively isolate lesions from the background. For classification, convolutional neural networks (CNNs) and their variants—including DenseNet, ResNet, U-Net, Inception, EfficientNet, AlexNet, VGG, NASNet, and GAN-based models—have shown high accuracy in distinguishing benign and malignant lesions. Performance evaluation typically employs metrics such as accuracy, precision, recall, and F1-score to assess model effectiveness. The review highlights the importance of robust datasets, effective preprocessing, and advanced

deep learning architectures, while noting the ongoing need for improved model interpretability, generalization, and clinical integration of computer-aided diagnosis systems for skin lesion analysis.

[7] Thwin and Park (2024), in their study titled "**Skin Lesion Classification Using a Deep Ensemble Model**", propose a novel approach for automated skin lesion classification by employing a deep learning ensemble that combines VGG16, ResNet-50, and Inception-V3 architectures. The study addresses the challenges of early skin cancer detection, particularly melanoma, where traditional visual diagnosis is often subjective. Using dermoscopic images from the ISIC 2018 and HAM10000 datasets, the authors apply oversampling to address class imbalance and use data preprocessing and augmentation to enhance model performance. Each CNN model is fine-tuned with transfer learning, utilizing pre-trained ImageNet weights and additional fully connected layers optimized for skin lesion classification. The ensemble integrates predictions from all three models through a weighted averaging method, leveraging their complementary strengths. Experimental results demonstrate that the ensemble significantly outperforms individual models, achieving up to 97% accuracy on ISIC 2018 and 96% on HAM10000, with superior precision, recall, and F1-scores. The approach is robust against class imbalance and highlights the clinical potential of ensemble deep learning models as decision support tools. The study also addresses practical challenges for clinical deployment, including data privacy, model interpretability, integration into clinical workflows, and ethical considerations, suggesting future directions such as incorporating multimodal data, advanced augmentation techniques like GANs, and real-world validation to enhance reliability and adoption in healthcare settings.

[8] Ji et al. (2024), in their study titled "**EFAM-Net: A Multi-Class Skin Lesion Classification Model Utilizing Enhanced Feature Fusion and Attention Mechanisms**", propose a novel deep learning framework for accurate multi-class skin lesion classification. EFAM-Net builds upon the ConvNeXt architecture and integrates three innovative components: the Attention Residual Learning ConvNeXt (ARLC) block for extracting low-level features with residual attention, the Parallel ConvNeXt (PCNXt) block for capturing global and complex features, and the Multi-scale Efficient Attention Feature Fusion (MEAFF) block to effectively fuse multi-level features. The model was trained and evaluated on ISIC 2019, HAM10000, and a private curated dataset, with preprocessing steps including resizing, data augmentation, and normalization. EFAM-Net achieved state-of-the-art performance, with accuracy rates of 92.3% on ISIC 2019, 93.95% on HAM10000, and 94.31% on the private dataset, outperforming baseline models such as ConvNeXt, ResNet-101, and EfficientNet. Grad-CAM visualizations demonstrated the model's superior ability to focus on lesion areas, enhancing interpretability. The design emphasizes robustness, efficient attention mechanisms, and multi-scale feature fusion, enabling high classification accuracy with minimal parameter increase. EFAM-Net provides a promising AI-driven tool for early detection and diagnosis of skin cancers, offering potential improvements in clinical decision-making and patient outcomes.

References

[1] Mirikharaji, Zahra, et al. "A survey on deep learning for skin lesion segmentation." Medical Image Analysis 88 (2023): 102863.

[2] Naqvi, Maryam, et al. "Skin cancer detection using deep learning—a review." Diagnostics 13.11 (2023): 1911.

[3] Al-Waisy, Alaa S., et al. "A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images." Scientific Reports 15.1 (2025): 31234.

[4] Hosny, Khalid M., et al. "Explainable deep inherent learning for multi-classes skin lesion classification." Applied Soft Computing 159 (2024): 111624.

[5] Sulthana, Razia, et al. "A novel end-to-end deep convolutional neural network based skin lesion classification framework." Expert Systems with Applications 246 (2024): 123056.

[6] Nugroho, Arief Kelik, et al. "Image dermoscopy skin lesion classification using deep learning method: systematic literature review." Bulletin of Electrical Engineering and Informatics 13.2 (2024): 1042-1049.

[7] Thwin, Su Myat, and Hyun-Seok Park. "Skin lesion classification using a deep ensemble model." Applied Sciences 14.13 (2024): 5599.

[8] EFAM-Net: A Multi-Class Skin Lesion Classification Model Utilizing Enhanced Feature Fusion and Attention Mechanisms