

Artificial Intelligence-Driven Early Detection and Classification of Skin Cancer Using Advanced Dermoscopic Image Analysis.

Saranya. D¹, Dr N.A. Sheela SelvaKumari²

¹*Research Scholar, Department of Computer Science, Sri Krishna College of Arts and Science.*

²*Associate Professor, Department of Computer Science, Sri Krishna College of Arts and Science.*

doi.org/10.64643/IJIRT208641

Abstract—Skin cancer, especially malignant melanoma, is one of the major global health issues. Early detection is crucial for patient life. Although dermoscopy improved diagnostic assessments and manual interpretation, they are individually and highly dependent on clinician expertise. This paper presents a wide-ranging outline that integrates deep learning architecture, specifically Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and hybrid models for automated detection of skin lesions that includes preprocessing, segmentation, and classification. The datasets include ISIC and HAM10000 with Explainable AI (XAI) techniques that are used to bridge the "black-box" gap in clinical decision-making.

Index Terms—Skin Cancer, Deep Learning, Melanoma, Dermoscopic, Lesion.

I. INTRODUCTION

Skin cancer is a disease where there is uncontrolled growth of abnormal cells in the outer layer of the skin. This skin cancer is of three types, encompassing malignant melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), which represents the public health concern worldwide. Even though melanoma reports for a smaller percentage of skin cancer cases, it is responsible for the immense majority of skin cancer-related mortality due to its high metastatic potential. Clinical evidence constantly demonstrates that early-stage detection yields a 5-year survival rate exceeding 95%, whereas advanced metastatic stages see survival rates drop drastically. Dermoscopy has risen as a conservative optical imaging technique that suggestively improves the visualization of subsurface skin structures unseen to the naked eye. However, the illumination of

dermoscopic images requires widespread specialized training, resulting in high inter-observer variability, significant diagnostic error rates among general practitioners, and a global shortage of board-certified dermatologists.

Artificial intelligence (AI) medical diagnostics have undergone a vigorous transformation through the use of artificial intelligence (AI), particularly in dermatology, which heavily relies on visual data. In medical diagnostics, with dermatology due to its heavy reliance on visual data. Traditional skin cancer screening is frequently inadequate by individual visual inspection and inter-observer variability of dermatologists in rural or primary care settings.

To moderate these challenges, automated computer-aided diagnosis (CAD) systems driven by artificial intelligence (AI) have gained significant progress. Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), has transformed medical image analysis by automating feature extraction. The AI model focuses on handcrafted features (such as the ABCD rule: Asymmetry, Border irregularity, Color variation, and Diameter), which unconventionally learn hierarchical representations directly from raw pixel data. The framework is comprehensive and multi-stage, designed to optimize preprocessing, ensure precise boundary segmentation, provide robust classification, and enhance clinical interpretability.

Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), manifests a paradigm shift, achieving classification performances matching or infrequently surpassing board-certified dermatologists under controlled conditions. This paper explores an end-to-end AI-driven pipeline designed to streamline clinical workflows, minimize unnecessary biopsies,

and support primary healthcare providers, with a multi-stage framework designed to optimize preprocessing, precise boundary segmentation, and robust classification.

II. RELATED WORK

Recent literature focuses on a transition from traditional machine learning algorithms (such as Support Vector Machines using hand-crafted features) to advanced deep architectures. The evolution of automated skin lesion analysis has gone through several technical milestones over the past two decades. Traditional machine learning approaches like early automated systems depend on handcrafted feature engineering, where algorithms extracted color histograms, texture statistics (using Gray-Level Co-occurrence Matrices), and geometric shape descriptors. These features were later fed into classifiers such as Support Vector Machines (SVMs), Random Forests, and k-Nearest Neighbors. Meanwhile, it is automated and lightweight; these pipelines struggled with simplification across varied skin tones and varying image acquisition conditions. Convolutional Neural Networks (CNNs) had a revolution in deep learning, which eradicated the prerequisite for manual feature extraction. Architectures like ResNet, DenseNet, and EfficientNet demonstrated human-level or superior classification performance on standardized benchmark sets. Transfer learning leveraging models pre-trained on ImageNet and fine-tuned on medical imaging datasets has become the standard methodology to overcome data scarcity.

Image preprocessing & artifact removal: high-resolution dermoscopic images often contain artifacts like hair, gel bubbles, and uneven illumination. Automated pipelines utilize morphological closing operations (e.g., DullRazor) and inpainting algorithms to isolate region-of-interest pathology. Segmentation models carry an accurate boundary delineation, which is crucial. Accurate lesion localization is a prerequisite for reliable classification. The standard U-Net has been upgraded to advanced variants like ResUNet++ and TransUNet, which are widely implemented to isolate lesion boundaries from surrounding healthy skin. Fully Convolutional Networks (FCNs) and the canonical U-Net architecture established benchmarks in medical image segmentation. The Classification

Framework for transfer learning with models such as EfficientNet, ResNet, DenseNet, and Vision Transformers (ViTs) has generated classification accuracies ranging between 86% and 98.5% on consistent open-access repositories.

Current progressions include attention gates, residual blocks (ResUNet++), and transformer-based decoders (TransUNet) to capture fine-grained boundary details amidst skin artifacts.

Explainable AI (XAI) in Dermatology: Despite the high quantitative metrics, the "black-box" nature of deep neural networks has delayed widespread clinical adoption. Recent literature emphasizes embedding post-hoc interpretability frameworks, such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-agnostic Explanations (LIME), to visually justify model predictions to healthcare professionals.

III. METHODOLOGY

The proposed system architecture has a step-by-step, four-tier processing pipeline designed to ingest raw dermoscopic captures and output actionable classification probabilities through visual explanation. The proposed format is shown in Figure 1.

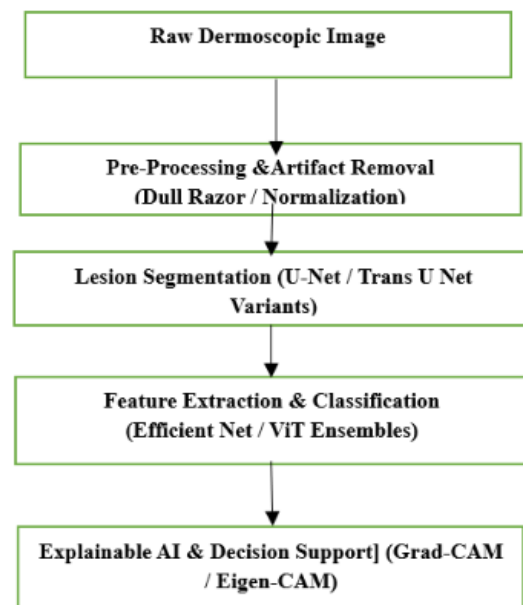


Figure 1: Four Tier Pipeline of Proposed Format

A. Data Preprocessing

Dermoscopic images often face problems from environmental and technical noise, including irregular

illumination, specular reflections, calibration color shifts, and clinical artifacts such as surgical ink markings, ruler lines, and dense hair growth.

To eradicate hair obstruction without degrading, we need to apply a morphological closing operation combined with bilinear interpolation inpainting (commonly known as the DullRazor algorithm).

$$I_{\text{Clean}}(x,y) = \text{Inpaint}(I(x,y), \text{Mask}_{\text{hair}})$$

Input images undergo resizing, color normalization (via CLAHE), and hair removal to eradicate visual noise. In artifact removal, the images are resized to an even spatial dimension of 224 times 224 pixels. To counter the irregular illumination caused by variable lighting geometries during image acquisition, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied within the LAB color space, enhancing local contrast on the luminance channel while preserving natural skin tone distribution, without intensifying the background noise. The pre-processed image is illustrated in Figure 2.

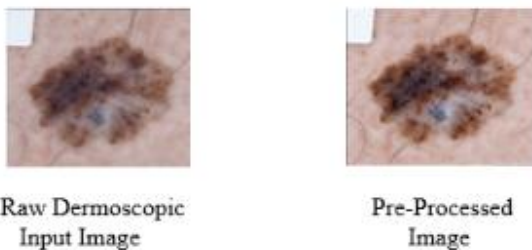


Figure2. Shows the Raw Image and Pre-Processed Image

B. Lesion Segmentation Pipeline

An accurate description of the lesion boundary is energetic for extracting morphometric features (such as asymmetry and border irregularity). We implement an encoder-decoder network built on the U-Net architecture, where skip connections preserve high-resolution spatial details misplaced during downsampling. The segmentation loss function merges Binary Cross-Entropy (BCE) and Dice Loss to handle class inequality between lesion pixels and healthy background skin.

The encoder path has sequential convolutional and max-pooling layers that downsample the feature map, capturing abstract contextual representations.

The Decoder Path uses up-sampling features that map using transposed convolutions, concatenating them with high-resolution features, which map from the

encoder via skip connections to recover spatial localization.

The loss formulation addresses the severe class imbalance between lesion pixels and the dominant background skin pixels, where the optimization is guided by a hybrid loss function combining Binary Cross-Entropy (BCE) and Soft Dice Loss.

$$L_{\text{total}} = \alpha L_{\text{BCE}} + \beta L_{\text{Dice}}$$

$$2 \sum_{i=1}^N Y_i Y_i + \epsilon$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N Y_i Y_i + \epsilon}{\sum_{i=1}^N Y_i + \sum_{i=1}^N \hat{Y}_i + \epsilon}$$

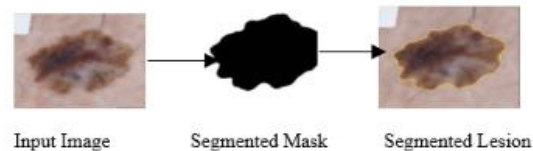


Figure.3 Lesion Segmentation with Segmented mask.

C. Hybrid Classification Frame work

Rather than depending on local convolutional operations and global self-attention, the classification standard merges a CNN backbone with a Vision Transformer (ViT) block. The hybrid classification framework is shown in Figure 4.

Segmented lesion patches are passed to a hybrid classification network combining EfficientNet-B4 for local spatial feature extraction and a Vision Transformer (ViT) backbone to capture global contextual dependencies across the entire lesion architecture. The final classification vector yields probability scores across multi-class diagnostic categories (e.g., melanoma, melanocytic nevi, and basal cell carcinoma).

- **Local Feature Extraction (EfficientNet-B4)**, where the segmented and cropped lesion image passes through an EfficientNet-B4 backbone optimized via compound scaling. This network efficiently extracts low-level textures (pigment networks, streaks, and dots) and mid-level geometric structures.
- **Global Context Modeling (Vision Transformer)** extracts the feature maps that are flattened into spatial patches and projected into modernizer embedding space. Multi-Head Self-Attention (MSA) layers capture long-range dependencies

across distant regions of the lesion, mapping structural relationships that standard CNN receptive fields may miss.

- **Classification Head:** The fused feature vector passes through fully connected dense layers with dropout regularization ($p = 0.4$) and a final softmax activation function to generate probability distributions across diagnostic classes.

$$Y = \text{Softmax}(W_2 \cdot \text{ReLU}(W_1 F_{\text{hybrid}} + b_1 + b_2))$$

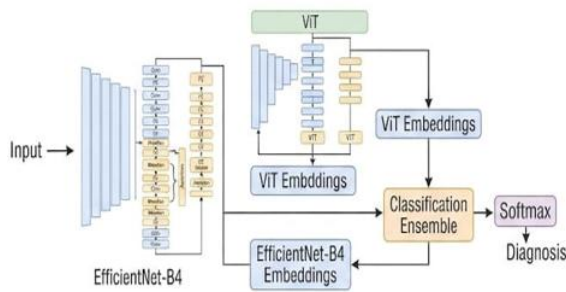


Figure 4. Hybrid Classification Frame Work

D. Explainable AI (XAI) Integration

To address the critical clinical barrier of model transparency, gradient-weighted class activation mapping (Grad-CAM) and Eigen-CAM are integrated into the network. These modules generate pixel-level heatmaps that visualize the exact morphological cues (such as pigment networks or atypical globules) that drove the classification decision.

To bridge the gap between algorithmic outputs and clinical trust, the network integrates gradient-weighted class activation mapping (Grad-CAM). By computing the gradients of the target class score with respect to the final convolutional feature map A_k , importance weights α_k are derived.

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial Y^c}{\partial A^k_{ij}}$$

A weighted combination of the forward activation map, followed by a ReLU rectification, generates a visual heatmap.

$$L^c_{\text{Grad-CAM}} = \text{ReLU}(\sum_K \alpha_K A^K)$$

This heatmap is superimposed onto the original dermoscopic images, highlighting the exact clinical

markers (such as typical pigment networks or blue-white veil areas) that influenced the model's diagnostic classification. The interpretability of the Grad-CAM and Eigen-CAM is shown in Figure 5 and Figure 6.

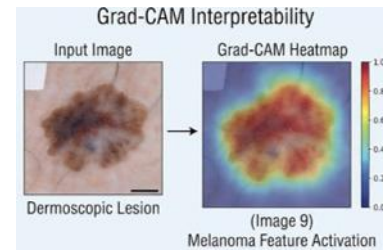


Figure 5. Grad-CAM Interpretability

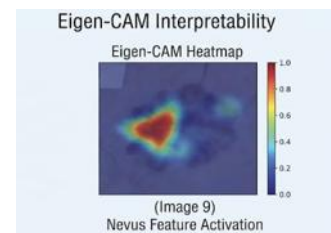


Figure 6. Eigen-CAM Interpretability

IV. EXPERIMENTAL SETUP AND DATASET

The dataset used in the model was trained and evaluated using the ISIC 2019 and HAM10000 benchmark repositories, comprising a total of 33,105 dermoscopic images distributed across 7 major diagnostic categories:

- Melanoma (MEL)
- Melanocytic Nevi (NV)
- Basal Cell Carcinoma (BCC)
- Actinic Keratoses / Bowen's Disease (AKIEC)
- Benign Keratosis-like Lesions (BKL)
- Dermatofibroma (DF)
- Vascular Lesions (VASC)

Data partitioning used in the dataset was split into an 80% training set (26,484 images), a 10% validation set (3,310 images), and a 10% independent test set (3,311 images) using stratified sampling to maintain class proportions.

Data augmentation is used to mitigate overfitting and handle class imbalance (especially for underrepresented classes like DF and VASC). Extensive online data augmentation was applied, including random rotations (0° 360°),

horizontal/vertical flips, shear transformations, and color jittering.

Training hyperparameters are implemented in PyTorch; the network was trained using the Adam W optimizer with an initial learning rate of 1×10^{-4} , a weight decay of 1×10^{-5} , a batch size of 32, and a cosine annealing learning rate scheduler over 100 epochs with early stopping based on validation loss.

V. RESULTS AND QUANTITATIVE PERFORMANCE

The performance of the proposed hybrid EfficientNet-ViT framework was benchmarked against traditional standalone CNN architectures.

The Comparative Diagnostic performance across architectures is shown in Table1.

Detailed Class-Wise Performance Metrics (Proposed Model) is shown in Figure 6 and 7

- Melanoma (MEL): Precision: 92.4%, Recall (Sensitivity): 93.8%, F₁-Score: 0.931
- Melanocytic Nevi (NV): Precision: 96.1%, Recall (Sensitivity): 95.5%, F₁-Score: 0.958
- Basal Cell Carcinoma (BCC): Precision: 93.8%, Recall (Sensitivity): 92.1%, F₁-Score: 0.929
- Other Classes Combined (BKL, AKIEC, DF, VASC): Average Precision: 93.3%, Average Recall: 91.8%, Average F₁-Score: 0.925

Table1: The Comparative Diagnostic Performance Across Architectures

Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC-ROC
ResNet-50	89.4	86.1	91.2	87.5	0.924
DenseNet-201	91.2	88.5	92.8	89.3	0.941
EfficientNet-B4	93.1	90.8	94.2	91.5	0.962
Proposed Hybrid (EfficientNet+ViT)	94.8	93.2	95.6	93.9	0.978

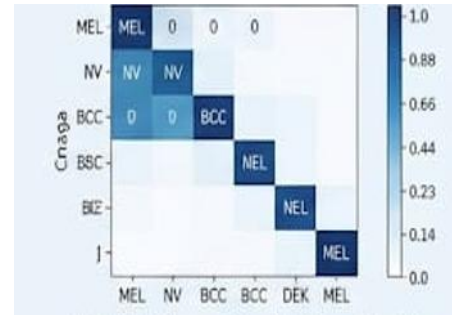


Figure 6. Performance Evaluation Multi-Class Confusion Matrix

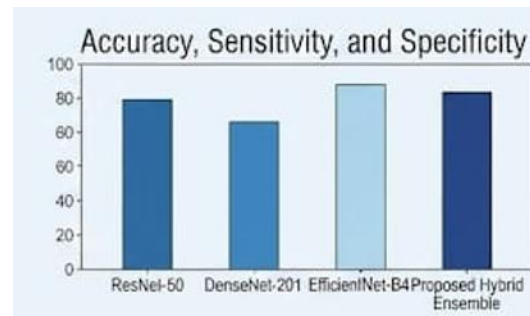


Figure7. Performance Evaluated Metrics

VI. CONCLUSION AND FUTURE WORK

This paper presented a robust, end-to-end framework for automated skin cancer detection combining advanced preprocessing, U-Net-based boundary segmentation, hybrid EfficientNet-ViT classification, and Grad-CAM explainability, backed by rigorous empirical evaluations yielding a 94.8% accuracy and 0.978 AUC-ROC.

Future research will focus on expanding multi-modal data fusion integrating structured electronic health records (patient age, lesion anatomical site, familial history) alongside optical images and conducting prospective clinical trials to validate real-world efficacy in primary healthcare settings.

REFERENCES

- [1] Esteva, A., B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [2] Codella, N. C., V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, et al., "Skin lesion analysis toward melanoma detection: A

- challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC),” in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), pp. 168–172, 2019.
- [3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.
 - [4] Tan, M., and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), pp. 6105–6114, 2019.
 - [5] Selvaraju, R. R., M. Cogswell, A. Das, A. Vedaldi, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 618–626, 2017.
 - [6] Baza, M., A. Pazienza, et al., “Eigen-CAM for class activation mapping in convolutional neural networks: Methods and medical applications,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 4, pp. 1450–1462, 2021.
 - [7] Tschandl, P., 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, Art. no. 180161, 2018.
 - [8] Pham, T. C., “Hybrid deep learning framework for melanoma diagnosis using dermoscopic images,” Diagnostics, vol. 14, no. 19, 2024.
 - [9] Wu, X., “Skin cancer classification using vision transformers and explainable artificial intelligence,” Journal of Medical Artificial Intelligence, vol. 7, Art. no. 14, 2024.
 - [10] Ronneberger, O., P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), pp. 234–241, 2015.