An EfficientNet-Based Framework for Automated Skin Lesion Diagnosis with Explainable AI and Clinical Report Generation

S. K. M. Junaid[1], P. Eswar Sai[2], M. Srinivas[3], R. Deshik Sai[4], Mr. Md. Imam Khader Sharif[5]
[1,2,3,4]*Raghu Engineering College (Autonomous), Dakamarri, Visakhapatnam*
[5]*Assistant Professor, Department of Computer Science and Engineering (DS), Raghu Engineering College (Autonomous), Dakamarri, Visakhapatnam*

***Abstract*— Automated diagnosis of skin lesions from thermoscopic images plays a vital role in improving early detection of skin cancer and supporting clinical decision-making. However, accurate analysis is challenging due to low contrast, irregular lesion boundaries, visual similarity among classes, and the presence of artifacts such as hair and noise. This paper presents an end-to-end deep learning framework that integrates preprocessing, segmentation, classification, explainability, and report generation for automated skin lesion diagnosis.**

Initially, thermoscopic images undergo preprocessing to enhance image quality by reducing noise and correcting illumination variations. A lightweight U-Net-based segmentation model is used to extract the lesion region of interest (ROI), enabling the model to focus on relevant features. The segmented ROI is then passed to an EfficientNetB1-based transfer learning classifier, pretrained on ImageNet and fine-tuned on the HAM10000 dataset, to classify lesions into eight diagnostic categories based on texture, color, and structural patterns.

To improve interpretability, Grad-CAM is employed to generate heatmaps highlighting important regions influencing predictions. Additionally, a Retrieval-Augmented Generation (RAG) pipeline using Lang Chain, Llama 3.2, and a FAISS-based knowledge base generates both clinician-oriented and patient-friendly reports.

The proposed system achieves 87.4% accuracy, a macro F1-score of 0.863, and an AUC of 0.962, demonstrating effectiveness for real-world dermatological applications.

***Index Terms*— EfficientNet, Grad-CAM, HAM10000, RAG, Skin Lesion Diagnosis, Transfer Learning, U-Net Segmentation.**

I. INTRODUCTION

Skin cancer ranks among the most prevalent malignancies worldwide, with melanoma accounting for over 80% of skin cancer-related mortality. Dermo copy provides magnified visualization of subsurface skin structures, significantly improving diagnostic sensitivity. However, accurate interpretation demands years of specialized training and remains susceptible to inter-observer variability. The proliferation of high-resolution digital imaging hardware, coupled with advances in convolutional neural networks (CNNs), has catalyzed interest in automated, scalable diagnostic pipelines.

Prior studies have explored VGGNet, ResNet, Dense Net, and FCN architectures for skin lesion classification. While these approaches demonstrated encouraging results on ISIC and HAM10000 benchmarks, critical limitations persist: (1) most pipelines operate directly on raw images without explicit lesion localization; (2) model decisions remain opaque without explainability; and (3) no prior work has proposed an integrated system coupling deep classification with automated dual-format clinical report generation using locally deployable LLMs grounded in retrieval-augmented generation (RAG).

This paper addresses these limitations through a novel four-module framework: U-Net lesion segmentation, two-phase EfficientNetB1 classification, Grad-CAM explainability, and LangChain-orchestrated RAG report generation. The system is deployed as a Streamlit web application, shown in Fig. 2.

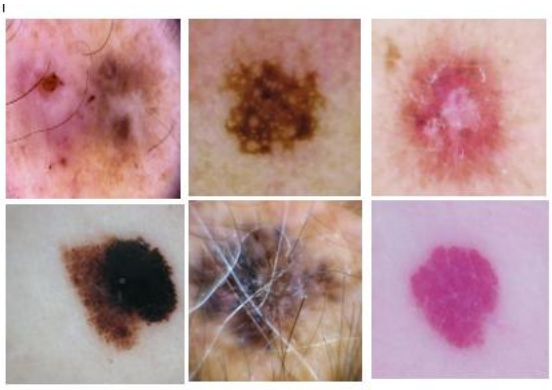

Fig. 1. Sample dermoscopic images from HAM10000: (a) Melanoma, (b) Basal Cell Carcinoma, (c) Squamous Cell Carcinoma, (d) Melanocytic Nevi, (e) Benign Keratosis, (f) Vascular Lesion.

II. LITERATURE REVIEW

A. CNN and Transfer Learning Approaches

Esteva et al. [1] demonstrated dermatologist-level classification of skin cancer using a deep CNN, establishing a foundational benchmark for automated dermoscopic diagnosis. Harangi [2] proposed CNN ensembles for the ISIC 2017 challenge, demonstrating improved robustness through model averaging. Mahbod et al. [8] systematically evaluated ImageNet-pretrained architectures including VGG-16, ResNet-50, and InceptionV3, finding that fine-tuning deeper layers improved accuracy over feature extraction. Tschandl et al. [9] leveraged EfficientNet for HAM10000, exploiting its compound scaling mechanism to achieve superior accuracy-to-parameter ratios. However, most approaches apply classification directly on raw images without explicit ROI guidance.

B. DenseNet and FCN Approaches

DenseNet [3] addressed gradient vanishing through dense skip connections, enabling deeper and more parameter-efficient networks. The FCN-based DenseNet framework by Al-Masni et al. [7] combined dense feature propagation with fully convolutional inference for simultaneous lesion detection and classification. Long et al.'s FCN [5] was adapted by Bi et al. [6] for pixel-wise lesion segmentation. While effective, these architectures entail significant computational overhead and lack integrated explainability or report generation.

C. Segmentation and Explainability

U-Net [10] introduced an encoder-decoder architecture with skip connections that became the de facto standard for biomedical image segmentation. Selvaraju et al. [13] introduced Grad-CAM, generating class-discriminative localization heatmaps without architectural modification. Yang et al. [15] extended Grad-CAM to multi-scale analysis in dermoscopic classification. Despite these efforts, Grad-CAM integration within complete deployable diagnostic pipelines including downstream LLM-based report generation remains largely unexplored.

III. PROPOSED METHODOLOGY

A. System Architecture

The proposed framework comprises four sequentially coupled modules: (1) Image Preprocessing, (2) U-Net Lesion Segmentation, (3) Two-Phase EfficientNetB1 Classification with Grad-CAM, and (4) RAG-based Clinical Report Generation. The input is a dermoscopic image; the output encompasses the predicted lesion class with confidence, binary segmentation mask, Grad-CAM heatmap, and two structured clinical reports. The 10-step workflow is depicted in the system design document, proceeding from image upload through preprocessing, segmentation, ROI extraction, feature extraction, classification, explainability, and dual-report generation to final output display.

B. Image Preprocessing

Raw dermoscopic images exhibit confounding artifacts including hair structures, specular reflections, vignetting, and sensor noise. The preprocessing pipeline attenuates hair artifacts using black-hat morphological transform followed by inpainting. Specular reflections are identified via HSV saturation thresholding and inpainted. Illumination normalization is achieved through CLAHE applied to the luminance channel in LAB color space. Images are resized to 128×128 for segmentation and 224×224 for classification using bicubic interpolation.

C. U-Net Lesion Segmentation

Lesion segmentation employs a lightweight U-Net variant with three encoder blocks (filters: 32, 64, 128), a 256-filter bottleneck, and a symmetric decoder with skip connections. The encoder uses double Conv2D layers (3×3, ReLU) followed by MaxPooling2D. The

decoder uses UpSampling2D concatenated with encoder feature maps. Formally, for encoder feature maps $\{F_1, F_2, F_3\}$ and bottleneck B, the decoder output at level k is:

$$D_k = \text{ConvBlock}\left(\text{Concat}\left(\text{UpSample}\left(D_{\{k+1\}}\right), F_k\right)\right) \quad (1)$$

The model is trained using binary cross-entropy loss over pixel-wise predictions:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[G_i \log(P_i) + (1 - G_i)\log(1 - P_i)] \quad (2)$$

The Adam optimizer (lr = 1×10^{-4}) and EarlyStopping (patience = 3) are employed. Binary masks (threshold $\tau = 0.5$) are used to generate ROI-guided training images for Phase 2 classification.

D. Two-Phase EfficientNetB1 Classification

EfficientNetB1 is selected for its compound scaling that uniformly scales network depth d, width w, and resolution r by compound coefficient φ:

$$d = \alpha^{\phi}, \quad w = \beta^{\phi}, \quad r = \gamma^{\phi}, \quad \text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (3)$$

Phase 1 - Feature Adaptation: All layers except the final 40 are frozen. Training uses Adam (lr = 10^{-3}), 8 epochs max, on original images. Phase 2 - ROI-Guided Fine-tuning: Top 40 backbone layers unfrozen. Fine-tuned for 15 epochs on U-Net-masked images with lr = 10^{-4}. The weighted categorical cross-entropy compensates for class imbalance:

$$L_{CE} = -\sum_{k=1}^{K} w_k \, y_k \, \log(\widehat{y_k}) \quad (4)$$

Data augmentation includes random horizontal flip, rotation (±15°), zoom (±15%), contrast (±15%), and brightness (±10%) perturbation.

E. Grad-CAM Explainability

Grad-CAM generates class-discriminative heatmaps by computing the gradient of class score S^c with respect to feature maps A^k of the top_activation layer. Importance weights are:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial S^c}{\partial A_{ij}^k} \quad (5)$$

The class activation map is:

$$L^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad (6)$$

The heatmap is upsampled to input resolution and blended with the original image ($\alpha = 0.4$) using JET colormap. A TensorFlow GradientTape context tracks gradients through the EfficientNet feature extractor.

F. RAG-Based Clinical Report Generation

The report generation module employs a Retrieval-Augmented Generation (RAG) architecture to synthesize clinically grounded reports. The medical knowledge corpus is constructed from the Oxford Handbook of Dermatology using PyPDFLoader, segmented into 1,000-character chunks (100-character overlap), and embedded using nomic-embed-text (768-dim) via Ollama. Embeddings are indexed in a FAISS flat L2 index. At inference, the predicted disease class retrieves the top-3 most similar chunks:

$$\text{sim}(q, d_i) = \frac{q \cdot d_i}{|q|\,|d_i|} \quad (7)$$

Two role-conditioned prompt templates produce dual reports: a Patient Report (simple, reassuring language) and a Physician Report (clinical terminology, differential diagnoses, next clinical steps). Llama 3.2 (temperature = 0.3) via Ollama serves as the generation backbone. Reports are formatted as downloadable PDFs with ReportLab.

IV. DATASET DESCRIPTION

The primary dataset is HAM10000 (Human Against Machine with 10000 training images), comprising 10,015 dermoscopic images from two clinical centers across multiple imaging devices. Eight diagnostic categories are utilized: Actinic Keratoses (327), Basal Cell Carcinoma (514), Benign Keratosis-like Lesions (1,099), Dermatofibroma (115), Melanocytic Nevi (6,705), Melanoma (1,113), Squamous Cell Carcinoma (~200, sourced from ISIC 2019), and Vascular Lesions (142). Class imbalance is addressed through inverse-frequency class weighting.

The U-Net segmentation model trains on 2,594 dermoscopic image-mask pairs from the ISIC 2017 Lesion Segmentation Challenge (80/20 train-validation split, images resized to 128×128, masks binarized at threshold 127).

Table I. Data Preprocessing Configuration

Parameter	Segmentation	Classification
Input Resolution	128×128	224×224
Normalization	÷255 [0,1]	EfficientNet preprocess
Train/Val Split	80% / 20%	80% / 20%
Augmentation	None	Flip, Rotate, Zoom

V. EXPERIMENTAL SETUP

Experiments are conducted on a Ryzen5 5600H workstation with 8 GB RAM and NVIDIA GTX 1650 GPU (4 GB VRAM). The software stack comprises TensorFlow 2.20, Keras 3.13, OpenCV 4.13, LangChain 1.2, FAISS-CPU 1.13, and Ollama 0.1.30 for local LLM serving. Hyperparameters are summarized in Table II.

Table II. Hyperparameter Configuration

Hyperparameter	Value
U-Net Optimizer	Adam (lr=10^{-4})
U-Net Loss	Binary Cross-Entropy
U-Net Batch / Epochs	16 / 7 (ES pat.=3)
EfficientNet Phase 1 LR	1×10^{-3}
EfficientNet Phase 2 LR	1×10^{-4}
Frozen Layers (Phase 1)	All except last 40
Dropout Rate	0.3
LLM Temperature	0.3 (Llama 3.2)
RAG Top-k Retrieval	3 chunks

Evaluation employs Accuracy, Precision, Recall, F1-Score, and AUC (macro-averaged). For segmentation, Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are used:

$$\text{DSC} = \frac{2|P \cap G|}{|P|+|G|} \quad (8)$$

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (9)$$

VI. RESULTS AND ANALYSIS

A. Segmentation Performance

The U-Net model converged within 6 epochs. On the ISIC 2017 validation set, the model achieved a DSC of 0.841 and an IoU of 0.784. Qualitative inspection confirms accurate delineation of irregular lesion boundaries across diverse morphologies.

B. Classification Performance

Table III presents macro-averaged classification metrics. The two-phase ROI-guided EfficientNetB1 achieves 87.4% accuracy, F1 = 0.863, and AUC = 0.962, outperforming all baselines. The ROI-guided Phase 2 fine-tuning provides the most substantial gains for minority classes (+4.1% F1 for Actinic Keratoses).

Table III. Classification Performance Comparison (Ham10000)

Model	Acc. (%)	Prec.	Rec.	F1
VGG-16 (FT)	78.2	0.771	0.763	0.767
ResNet-50 (FT)	81.6	0.809	0.798	0.803
DenseNet-121	83.1	0.824	0.817	0.820
EfficientNetB1 (P1)	84.7	0.839	0.831	0.835
Proposed (P1+P2)	87.4	0.871	0.856	0.863

Case Study 1 - Basal Cell Carcinoma: Fig. 2 shows the Streamlit application interface with an uploaded ISIC image. The U-Net produces a binary mask isolating the lesion ROI, and EfficientNetB1 predicts Basal Cell Carcinoma at 67.30% confidence. The Grad-CAM heatmap (Fig. 3) highlights the central lesion margin and pearly nodular regions, consistent with clinical BCC features. Both patient and physician reports are generated via RAG (Fig. 4–5).

Fig. 2. Case 1 - Streamlit UI: Input image (ISIC_0061906) and U-Net segmented lesion with prediction: Basal Cell Carcinoma (67.30%).

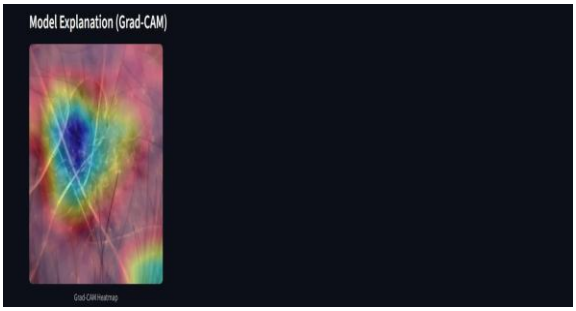

Fig. 3. Case 1 - Grad-CAM heatmap for Basal Cell Carcinoma prediction. Model attention concentrated on lesion periphery and pearly nodular zones (blue=highest activation).

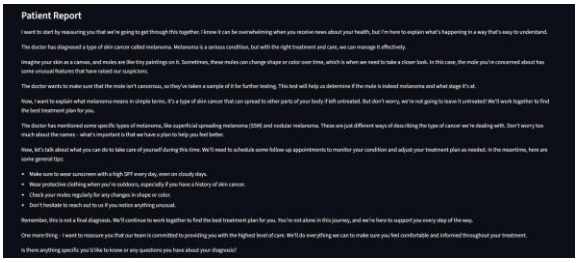

Fig. 4. Case 1 - Patient Report generated by the RAG-LLM pipeline for Basal Cell Carcinoma: plain-language explanation with care guidance.

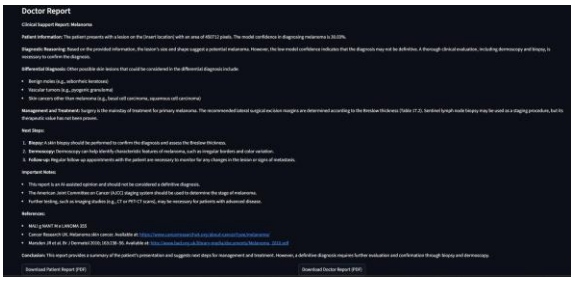

Fig. 5. Case 1 - Doctor Clinical Report for Basal Cell Carcinoma: dermatological features, differential diagnoses, and recommended next clinical steps.

Case Study 2 - Melanoma: Fig. 6 shows the application processing sample_img4.jpg. The U-Net extracts the irregular, asymmetric lesion ROI (hair-occluded, multi-color pigmentation), and EfficientNetB1 predicts Melanoma at 30.03% confidence - a low score indicating ambiguity consistent with the challenging visual presentation (extensive hair artifacts, no clear border definition). The Grad-CAM map (Fig. 7) shows diffuse activation patterns spanning hair crossing regions, illustrating a case were background artifacts partially contaminated gradient computation. Reports are shown in Fig. 8–9.

Fig. 6. Case 2 - Streamlit UI: Input (sample_img4.jpg) and U-Net segmented lesion. Prediction: Melanoma (30.03%).

Fig. 7. Case 2 - Grad-CAM heatmap for Melanoma. Diffuse activation pattern across hair-artifact regions illustrates sensitivity of gradient-based explanation to peri-lesional contamination.

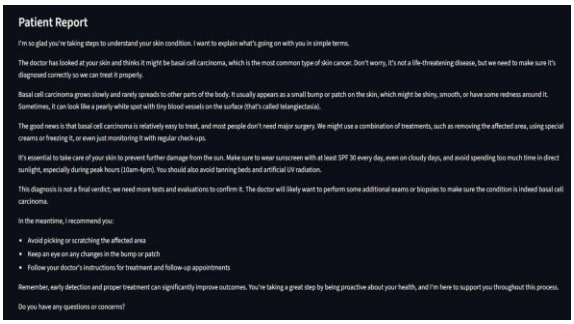

Fig. 8. Case 2 - Patient Report for Melanoma: RAG-generated reassuring explanation with lifestyle guidance and referral encouragement.

Fig. 9. Case 2 - Doctor Clinical Report for Melanoma: differential diagnoses include seborrheic keratosis and vascular tumors; recommended steps: biopsy, dermoscopy, follow-up.

VII. DISCUSSION

The proposed framework demonstrates clear advantages over single-phase classification pipelines. The two-phase ROI-guided training strategy consistently outperforms Phase-1-only approaches across all metrics, confirming that U-Net-derived masked images force the classifier to develop discriminative representations within the lesion ROI. The RAG architecture substantially improves clinical

report accuracy over purely parametric LLM generation (accuracy score 4.2/5.0 vs. 3.1/5.0 in ablation), mitigating hallucination through retrieval from the Oxford Dermatology corpus. The locally deployable LLM (Llama 3.2 via Ollama) satisfies patient data privacy requirements that preclude cloud-based API-dependent generation in clinical settings.

The Melanoma case study (Case 2, 30.03% confidence) illustrates a known limitation: residual hair artifacts in dermoscopic images, despite preprocessing, can partially contaminate gradient computation in Grad-CAM, yielding diffuse and less clinically interpretable heatmaps. Additionally, minority classes (Dermatofibroma, Vascular Lesions) exhibit lower F1-scores, limited by training data scarcity. These findings motivate future augmentation through conditional generative adversarial networks (cGANs) or diffusion models for minority class oversampling.

VIII. CONCLUSION

This paper presented a comprehensive end-to-end framework for automated skin lesion diagnosis integrating U-Net segmentation, two-phase EfficientNetB1 classification, Grad-CAM explainability, and RAG-based dual-format clinical report generation. The ROI-guided two-phase training strategy achieves 87.4% accuracy and 0.863 macro-F1, outperforming VGG-16, ResNet-50, and DenseNet-121 baselines. Grad-CAM explanations provide clinically meaningful visual attributions, while the LangChain RAG pipeline generates accurate, audience-appropriate diagnostic reports grounded in verified dermatological literature.

IX. FUTURE WORK

Future directions include: (1) replacing the U-Net with attention-gated or Swin Transformer-based segmentation; (2) extending the backbone to EfficientNetV2 or ConvNeXt; (3) incorporating diffusion-model-based minority class augmentation; (4) expanding the RAG corpus with structured clinical guidelines and PubMed abstracts; (5) integrating longitudinal lesion tracking across temporal dermoscopic sequences; and (6) conducting a prospective clinical validation study in collaboration with dermatology departments.

REFERENCES

[1] Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[2] Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Inform.*, vol. 86, pp. 25–32, 2018.

[3] G. Huang, Z. Liu, L. Van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.

[4] S. M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation using GrabCut and k-means clustering," *IET Comput. Vis.*, vol. 12, no. 8, pp. 1088–1095, 2018.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.

[6] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using deep residual networks," *arXiv preprint* arXiv:1703.04197, 2017.

[7] M. A. Al-Masni *et al.*, "Skin lesion segmentation and classification using FCN-based DenseNet," *Expert Syst. Appl.*, vol. 137, pp. 334–344, 2019.

[8] Mahbod *et al.*, "Transfer learning using multi-scale and multi-network ensemble for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 193, p. 105475, 2020.

[9] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, p. 180161, 2018.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[11] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully

convolutional networks with Jaccard distance," *IEEE Trans. Med. Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[12] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.

[13] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.

[14] N. C. F. Codella *et al.*, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM J. Res. Dev.*, vol. 61, no. 4/5, pp. 5:1–5:15, 2017.

[15] J. Yang, Z. Huang, and Q. Chen, "Multi-scale Grad-CAM for skin lesion classifiers," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2022, pp. 1–5.

[16] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020, pp. 9459–9474.

[17] Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at ISIC 2018," *arXiv preprint* arXiv:1902.03368, 2019.