ISSN 2723-6471

An Explainable Uncertainty-Aware Deep Learning Framework for Robust Skin Disease Diagnosis

R. Karthick Manoj[1,*], Aasha Nandhini S[2], C. Batumalai[3]

[1]*Dept. of Electrical and Electronics Engineering, AMET University, Chennai, India*

[2]*Dept. of Electronics and Communication Engineering, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, India*

[3]*Faculty of Data Science and IT, INTI International University, Nilai, Malaysia*

(Received: February 10, 2026; Revised: April 15, 2026; Accepted: July 15, 2026; Available online: August 18, 2026)

Abstract

Automated skin-disease diagnosis from clinical and dermoscopic images remains challenging because of illumination variability, acquisition artifacts, irregular lesion boundaries, limited feature diversity, and the low interpretability of conventional deep learning models. This study proposes EDLF-Skin, an integrated framework that combines Physio-Color Consistency preprocessing, Attentive Multi-scale U-Trans segmentation, Multi-Modal Feature Fusion, a Hybrid Ensemble with Uncertainty, and a Multi-level Explainability Interface. The preprocessing stage normalizes illumination and removes artifacts while preserving diagnostically relevant color and texture information. The segmentation module integrates convolutional feature extraction with transformer-based attention to improve lesion-boundary delineation. Spatial, frequency, and texture representations are subsequently fused through cross-attention and processed by complementary convolutional and transformer classifiers. Predictive uncertainty is estimated to identify low-confidence cases requiring expert review, while Grad-CAM, SHAP attribution, and textual rationales provide multilevel explanations. Experimental results show that the proposed segmentation model achieved a Dice coefficient of 0.956 and an intersection over union of 0.935. The complete classification framework obtained 96.87% accuracy, an F1-score of 0.953, an area under the receiver operating characteristic curve of 0.981, and an expected calibration error of 0.021. These findings indicate that EDLF-Skin improves segmentation accuracy, classification performance, probability calibration, and decision transparency within a unified diagnostic workflow.

Keywords: Skin Disease Diagnosis, Explainable Deep Learning, Lesion Segmentation, Multimodal Feature Fusion, Uncertainty Estimation, Grad-CAM, SHAP, Process Innovation

1. Introduction

Skin diseases represent a substantial global health burden and affect individuals across age groups, geographic regions, and socioeconomic conditions. Their clinical manifestations range from relatively mild inflammatory disorders to aggressive malignancies that require immediate intervention. Early and accurate identification is therefore essential for preventing disease progression, improving treatment outcomes, and reducing unnecessary diagnostic procedures. However, conventional diagnosis remains strongly dependent on visual examination, clinical experience, image quality, and access to trained dermatologists, which may vary considerably across healthcare settings [1]. These limitations have increased the demand for automated diagnostic systems capable of supporting consistent and timely skin-disease assessment [2].

Recent advances in deep learning have significantly improved the automated analysis of clinical and dermoscopic images. Convolutional neural networks can learn hierarchical representations of lesion color, shape, texture, and structural irregularities directly from image data, reducing dependence on manually designed features [3]. Transfer learning, image augmentation, and optimized network architectures have further improved classification performance, particularly when annotated medical datasets are limited [4]. Nevertheless, strong performance on benchmark datasets does not necessarily guarantee reliable operation under real clinical conditions, where images may differ in acquisition devices, illumination, resolution, patient skin tone, and lesion presentation. Variability in image acquisition remains a

***Corresponding author: R. Karthick Manoj (karthickmanoj.r@gmail.com)**
DOI: https://doi.org/10.47738/jads.v7i3.1465

major source of prediction instability. Shadows, uneven illumination, hair, ruler markings, reflections, and background artifacts can obscure diagnostically relevant regions and introduce patterns unrelated to the underlying disease [5], [6]. Standard resizing, normalization, and filtering procedures may reduce noise but can also alter subtle pigmentation and texture characteristics. Consequently, an effective diagnostic pipeline should normalize imaging conditions while preserving clinically meaningful information, including lesion boundaries, chromatic variation, asymmetry, and local texture patterns.

Accurate lesion segmentation is equally important because classification models may otherwise learn features from surrounding skin or image artifacts rather than from the lesion itself. Conventional segmentation approaches, including threshold-based methods and standard encoder–decoder networks, often struggle with low-contrast lesions, irregular boundaries, heterogeneous pigmentation, and substantial variation in lesion size [7]. Furthermore, many existing diagnostic systems rely predominantly on spatial representations extracted by a single network branch. Such representations may overlook complementary frequency and micro-texture information that can help distinguish visually similar disease categories. Integrating spatial, frequency-domain, and texture-based features may therefore provide a more discriminative and robust lesion representation.

Another unresolved issue is the limited transparency of high-performing deep neural networks. Their predictions are commonly produced without sufficient information about the image regions or characteristics that influenced the decision, restricting their clinical acceptability [5], [8]. Explainability approaches such as saliency mapping and SHAP-based attribution can provide insight into model behavior; however, they are frequently applied only after model development and are not integrated with segmentation, uncertainty estimation, or diagnostic reporting [9], [10]. Moreover, an explanation alone does not indicate whether a prediction is sufficiently reliable. A clinically oriented system should therefore provide interpretable evidence together with calibrated confidence and uncertainty estimates so that ambiguous cases can be referred for expert review.

To address these limitations, this study proposes EDLF-Skin, an integrated explainable deep learning framework for intelligent skin-disease diagnosis. The framework combines Physio-Color Consistency preprocessing, Attentive Multi-scale U-Trans lesion segmentation, Multi-Modal Feature Fusion, and a Hybrid Ensemble with Uncertainty to produce calibrated diagnostic predictions. A Multi-level Explainability Interface subsequently provides visual attention maps, feature-level attributions, and concise textual rationales. Unlike pipelines that optimize preprocessing, segmentation, classification, and explanation independently, EDLF-Skin integrates these components within a unified diagnostic workflow. The proposed design is intended to improve robustness against illumination and acquisition variability, strengthen lesion representation, quantify predictive uncertainty, and generate explanations that are more closely aligned with clinically relevant characteristics [9], [11].

2. Literature Review

Deep learning has become a dominant approach for automated skin-lesion analysis because of its ability to learn discriminative visual representations directly from clinical and dermoscopic images. Convolutional neural networks have been widely applied to identify lesion morphology, pigmentation, border irregularity, and texture patterns, while transfer learning and data augmentation have helped mitigate the limited availability of annotated medical images [3], [4]. Recent studies have reported strong classification performance across several benchmark datasets; however, their effectiveness often declines when models are evaluated on images acquired using different devices, illumination settings, demographic groups, or clinical environments [2], [6]. This limitation indicates that diagnostic performance should not be assessed solely through in-domain accuracy, but also through robustness to image-quality variation and domain shift.

Image preprocessing and lesion segmentation are therefore essential components of a reliable diagnostic pipeline. Color normalization and illumination correction have been used to reduce inconsistencies caused by shadows, exposure, camera characteristics, and skin-tone variation. More recent approaches employ learning-based and generative normalization methods to harmonize image appearance while retaining diagnostically relevant chromatic information [11]. For lesion localization, conventional U-Net architectures remain widely used, but their local convolutional operations may provide limited capability for modelling long-range spatial relationships. Hybrid CNN–Transformer

architectures have consequently been introduced to combine local feature extraction with global contextual modelling, producing more accurate segmentation for lesions with low contrast, irregular boundaries, or heterogeneous structures [12], [13]. Nevertheless, preprocessing and segmentation are frequently optimized as separate tasks, without evaluating how their outputs influence downstream classification reliability.

Feature representation is another critical factor in distinguishing skin conditions with visually overlapping characteristics. Most existing systems rely primarily on deep spatial features generated by a single CNN or vision-transformer backbone. Although these representations capture high-level lesion structures, they may not sufficiently encode subtle frequency changes and micro-texture variations associated with keratinization, pigmentation distribution, or surface irregularity [7]. Studies incorporating wavelet transformations, Gabor filters, handcrafted descriptors, and deep representations have demonstrated that spatial, frequency, and texture features can provide complementary diagnostic information [14], [15]. However, simple feature concatenation may preserve redundant information and assign equal importance to features with different diagnostic relevance. Attention-based fusion mechanisms are therefore required to adaptively determine the contribution of each representation.

Explainability and predictive reliability have also received increasing attention in computer-aided dermatology. Grad-CAM, SHAP, and related attribution methods can reveal whether model decisions are associated with clinically meaningful regions such as asymmetric structures, irregular borders, or uneven pigmentation [5], [8]. Evidence suggests that explanations aligned with dermatological criteria can improve physician trust and confidence in model-assisted melanoma assessment [9], [16]. Nevertheless, many studies implement explainability only as a post-hoc visualization and do not combine it with probability calibration or uncertainty estimation. As a result, a model may provide a plausible explanation even when its prediction is unreliable. Existing findings therefore support the need for an integrated framework that combines image normalization, precise segmentation, multimodal feature fusion, calibrated classification, uncertainty quantification, and multilevel explanation within a single diagnostic workflow.

3. Methodology

3.1. Study Design, Dataset Preparation, and Preprocessing

This study developed an end-to-end diagnostic framework, termed EDLF-Skin, for skin-disease classification from clinical and dermoscopic images. The experimental workflow consisted of five sequential stages: image preprocessing, lesion segmentation, multimodal feature extraction, uncertainty-aware classification, and prediction explanation. Two image collections were used to evaluate whether the framework could operate across different acquisition domains: DermSkin-2024, containing dermoscopic images from five diagnostic categories, and Clinical-SkinDB, containing clinical RGB photographs from three diagnostic categories. All images were converted to RGB format and resized to 224×224 pixels before model training. Because the original manuscript does not report the number of images, data sources, patient distribution, or train–validation–test allocation, these details must be added before submission. The dataset description presented in table 1.

Table 1. Dataset composition and experimental partition

Dataset	Image domain	Diagnostic classes	Input size
DermSkin-2024	Dermoscopic	Melanoma, nevus, basal cell carcinoma, actinic keratosis, benign lesion	(224 x 224)
Clinical-SkinDB	Clinical RGB	Melanoma, keratosis, nevus	(224 x 224)

Required verification: The official dataset names, public repositories, ethical status, and citation sources must be stated. If “DermSkin-2024” and “Clinical-SkinDB” are internally compiled datasets, the manuscript must explain the collection procedure, annotation process, patient consent, and institutional approval. The dataset should be divided at the patient level, rather than at the image level, to prevent images from the same patient appearing in both training and testing sets [17], [18], [19]. A recommended split is 70% training, 15% validation, and 15% testing, although the final manuscript must report the actual partition used. Stratified sampling should be applied to preserve the class distribution. During training, data augmentation should include random horizontal and vertical flipping, rotation within $\pm 20°$, scaling, translation, and controlled brightness and contrast perturbation. Augmentation must be applied exclusively to the training set.

Image preprocessing was performed using the proposed Physio-Color Consistency (PCC) module. The module estimates spatially varying illumination, suppresses acquisition artifacts, and preserves lesion-specific color and texture information [20]. Illumination normalization is expressed as:

$$I_c(x,y) = \frac{I(x,y)}{\hat{L}(x,y) + \epsilon} \tag{1}$$

$I(x,y)$ is the observed RGB image, $\hat{L}(x,y)$ is the estimated illumination field, $I_c(x,y)$ is the normalized image, and ϵ prevents numerical instability. The illumination field is estimated by minimizing

$$\hat{L} = \arg\min_{L} |I - R \odot L|_2^2 + \lambda_{\text{smooth}} |\nabla L|_2^2 \tag{2}$$

R represents image reflectance, $\odot$ denotes element-wise multiplication, and λ_{smooth} controls spatial smoothness. This formulation reduces illumination inconsistency while maintaining diagnostically relevant reflectance patterns. Color normalization was followed by bilateral filtering and morphological artifact detection to suppress hair, ruler markings, shadows, and specular reflections. Such normalization is important because acquisition variability can introduce domain-specific information unrelated to lesion pathology [6], [11], [21]. The existing preprocessing illustration can be retained but should be repositioned within this subsection. Figure 1 show the output from preprocessing stage.

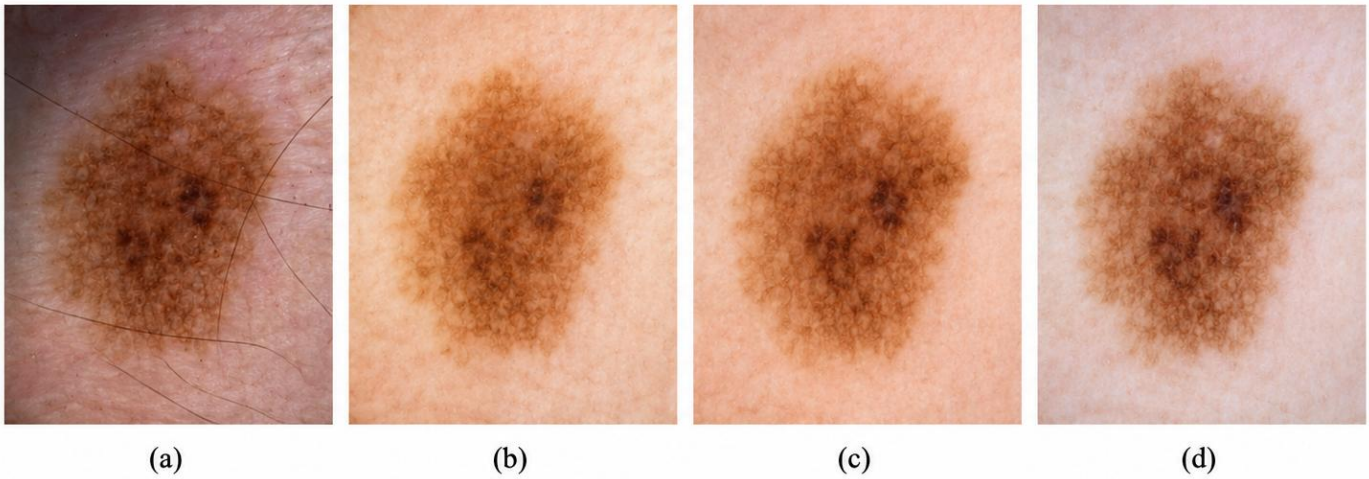

Figure 1. Output of the PCC preprocessing stage: (a) original image, (b) illumination-normalized image, (c) artifact mask or artifact-removed image, and (d) final color-normalized output.

3.1. Proposed EDLF-Skin Architecture

The proposed EDLF-Skin architecture integrates lesion preprocessing and segmentation, complementary multi-branch feature extraction, cross-attention-based feature fusion, and uncertainty-aware classification into a unified end-to-end diagnostic pipeline. As illustrated in figure 2, the input dermoscopic image is first processed using Physio-Color Consistency (PCC) preprocessing and segmented by AMU-Trans to obtain a lesion-masked image. The masked lesion is then analyzed through three complementary branches that capture spatial, frequency, and texture information using CNN, wavelet, and LBP/Gabor features, respectively. These heterogeneous representations are integrated through cross-attention fusion and subsequently processed by parallel CNN and Transformer classifiers. Their predictions are combined through an uncertainty-aware ensemble to produce the final disease label and confidence and uncertainty scores, while Grad-CAM and SHAP provide complementary visual and feature-level explanations.

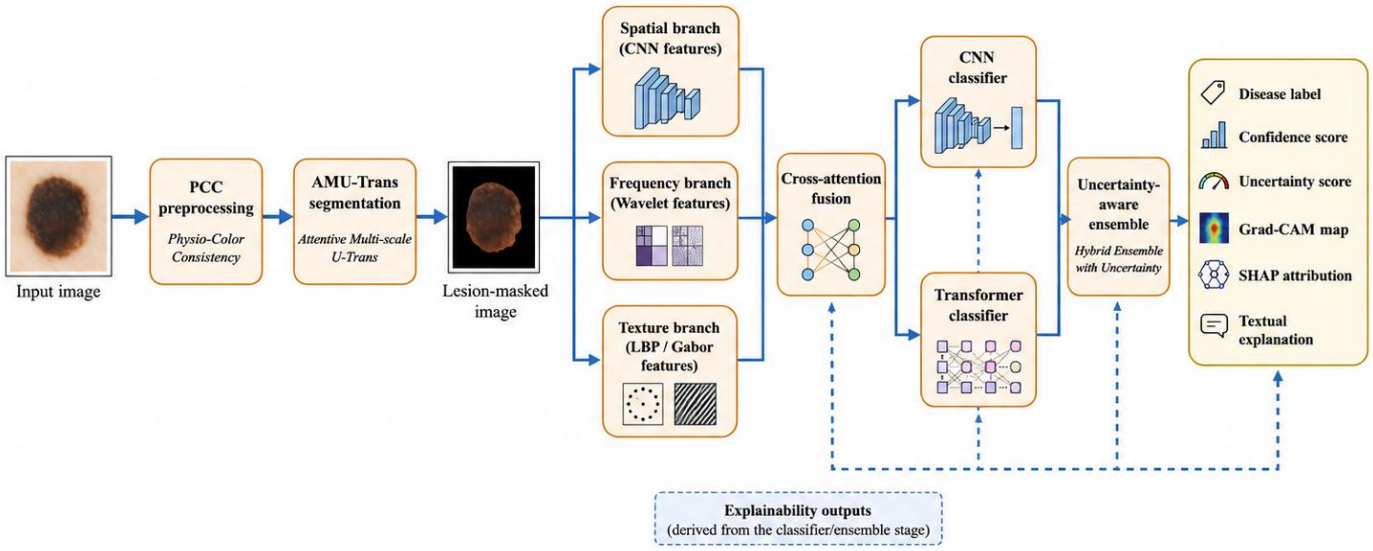

Figure 2. Overall EDLF-Skin architecture.

The preprocessed image I_c is first processed by the Attentive Multi-scale U-Trans (AMU-Trans) segmentation network. The network combines convolutional encoding for local morphology with transformer-based self-attention for long-range spatial dependencies. Its segmentation output is defined as

$$\widehat{M} = f_{\text{AMU-Trans}}(I_c; \theta_s), \qquad \hat{B} = g(\widehat{M}) \tag{3}$$

$\widehat{M}$ denotes the predicted lesion mask, $\hat{B}$ is the corresponding boundary-confidence map, and θ_s represents the trainable segmentation parameters. Transformer attention in the bottleneck is calculated as

$$Attention(Q, K, V) = softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{4}$$

Q, K, and V denote query, key, and value matrices, while d_k is the feature dimension used for scaling. The hybrid design is intended to preserve local boundary details while modelling broader relationships across the lesion region, consistent with recent CNN–Transformer segmentation developments [12]. Segmentation is optimized using a composite objective:

$$\mathcal{L}sg = \alpha\mathcal{L}_{\text{Dc}} + \beta\mathcal{L}\text{BE} + \gamma\mathcal{L}_{\text{budr}} \tag{5}$$

α, β, and γ control the contribution of overlap, pixel-wise classification, and boundary-preservation objectives. Dice loss is calculated as

$$\mathcal{L}_{\text{Dc}} = 1 - \frac{2\sum_i \widehat{M_i} M_i + \epsilon}{\sum_i \widehat{M_i} + \sum_i M_i + \epsilon} \tag{6}$$

and binary cross-entropy is defined as

$$\mathcal{L}\text{BE} = -\frac{1}{N}\sum_{i=1}^{N}\left[M_i \log\left(\widehat{M_i}\right) + (1 - M_i)\log\left(1 - \widehat{M_i}\right)\right]. \tag{7}$$

The original manuscript mentions a shape-prior autoencoder but does not explain its architecture, training data, or optimization procedure. Unless the authors can provide these details and corresponding experiments, the shape-prior claim should be removed and replaced by the explicitly defined boundary loss:

$$\mathcal{L}_{\text{budr}} = 1 - \frac{2\sum_i \widehat{B_i} B_i + \epsilon}{\sum_i \widehat{B_i} + \sum_i B_i + \epsilon} \tag{8}$$

B and $\hat{B}$ are the ground-truth and predicted lesion-boundary maps. The segmented lesion image is obtained through

$$I_{\text{lesion}} = \widehat{M} \odot I_c. \tag{9}$$

This image is passed to the Multi-Modal Feature Fusion (MMFF) module, which contains three parallel branches. The spatial branch uses a CNN backbone to learn semantic morphology, the frequency branch applies a trainable wavelet transformation, and the texture branch extracts micro-textural patterns using Gabor and local binary pattern representations. The branch outputs are expressed as

$$\begin{aligned} f_s &= F_{\text{CNN}}(I_{\text{lesion}}) \\ f_f &= W_{\psi}(I_{\text{lesion}}) \\ f_t &= F_{\text{texture}}(I_{\text{lesion}}) \end{aligned} \tag{10}$$

f_s, f_f, and f_t denote spatial, frequency, and texture features, respectively, and W_{ψ} represents the trainable wavelet transformation. These representations are integrated using cross-attention:

$$f_{\text{fused}} = F_{\text{cross-att}}(f_s, f_f, f_t) \tag{11}$$

This fusion mechanism allows the model to assign different importance to morphological, frequency, and micro-texture information rather than simply concatenating all features. The use of wavelet-based representations is supported by prior findings that frequency-domain features can capture lesion structures that may not be sufficiently represented by spatial features alone [14].

4. Results and Discussion

4.1. Preprocessing and Lesion Segmentation Performance

The proposed EDLF-Skin framework was evaluated on dermoscopic and clinical skin-image datasets to assess its segmentation accuracy, diagnostic performance, calibration, and interpretability. All experiments were implemented using PyTorch and executed on a workstation equipped with an Intel Core i9 processor, 64 GB of RAM, and an NVIDIA RTX 4090 GPU. Input images were resized to 224×224 pixels and processed using the Physio-Color Consistency module before lesion segmentation and classification.

Figure 3 presents representative outputs of the PCC preprocessing module. Illumination normalization reduced overexposed and underexposed regions, while morphological filtering and image restoration suppressed hair, ruler markings, and other acquisition artifacts. The resulting images retained the major chromatic and textural characteristics of the lesion while providing a more consistent appearance across samples. This normalization step reduced the influence of image-acquisition conditions and allowed the subsequent modules to focus more strongly on lesion-specific information.

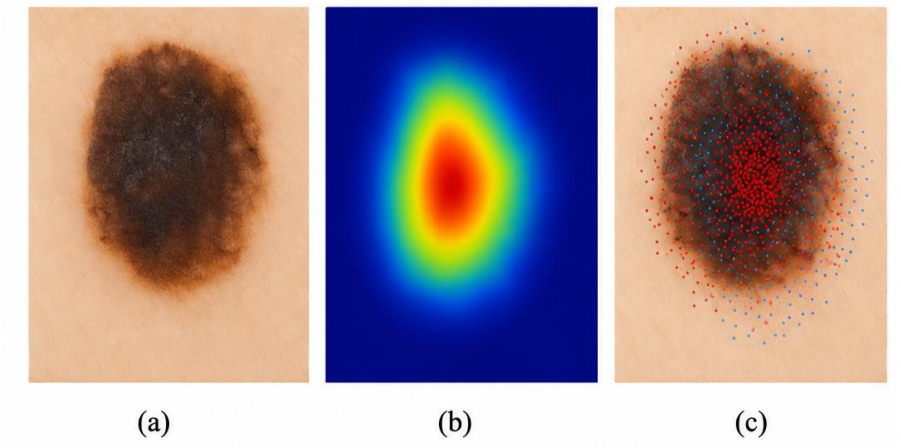

Figure 3. Representative outputs of the PCC preprocessing module: (a) original image, (b) illumination-normalized image, (c) artifact-removed image

The segmentation performance of AMU-Trans was compared with U-Net, Attention U-Net, and SegNet using Dice coefficient, intersection over union, precision, recall, and F1-score. As shown in table 2, AMU-Trans achieved the highest performance across all reported metrics, with a Dice coefficient of 0.956 and an IoU of 0.935. Compared with standard U-Net, the proposed model improved the Dice score by 0.045 and the IoU by 0.051. It also exceeded Attention U-Net by 0.029 in Dice and 0.032 in IoU.

Table 2. Segmentation performance comparison

Model	Dice	IoU	Precision	Recall	F1-score
U-Net	0.911	0.884	0.905	0.918	0.894
Attention U-Net	0.927	0.903	0.921	0.933	0.909
SegNet	0.892	0.868	0.887	0.896	0.875
Proposed AMU-Trans	0.956	0.935	0.951	0.958	0.947

The improvement can be attributed to the complementary roles of convolutional encoding and transformer-based attention (see figure 4). Convolutional blocks preserved fine-grained boundary information, whereas the transformer bottleneck captured broader spatial dependencies across lesion regions. Attention-guided skip connections further reduced the transfer of irrelevant background features to the decoder. Qualitatively, the predicted masks maintained irregular lesion shapes and produced fewer fragmented regions, particularly for low-contrast lesions.

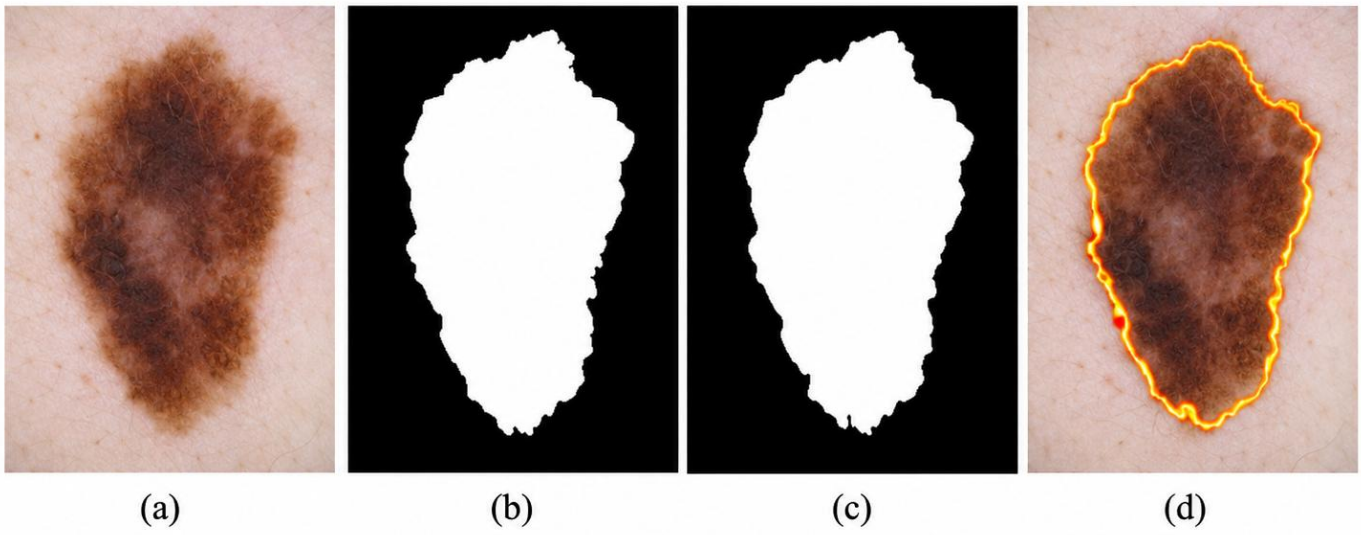

Figure 4. Representative AMU-Trans segmentation output: (a) input image, (b) ground-truth mask, (c) predicted lesion mask, and (d) boundary-confidence overlay.

4.2. Classification, Calibration, and Uncertainty Analysis

The Hybrid Ensemble with Uncertainty classifier was compared with EfficientNet-B7, DenseNet-201, and a Vision Transformer. Performance was evaluated using accuracy, F1-score, sensitivity, specificity, area under the receiver operating characteristic curve, and expected calibration error. Table 3 shows that the proposed HEU classifier achieved the best overall performance.

Table 3. Classification and calibration performance

Model	Accuracy (%)	F1-score	Sensitivity	Specificity	AUC	ECE
EfficientNet-B7	93.45	0.918	0.929	0.913	0.954	0.067
DenseNet-201	92.16	0.911	0.921	0.901	0.947	0.072
Vision Transformer	94.08	0.930	0.934	0.919	0.962	0.059
Proposed HEU	96.87	0.953	0.961	0.948	0.981	0.021

 ISSN 2723-6471

The proposed classifier improved accuracy by 2.79 percentage points over the Vision Transformer, which was the strongest standalone baseline. The F1-score increased from 0.930 to 0.953, indicating that the improvement was not limited to the majority classes. The sensitivity of 0.961 suggests that the model correctly detected a high proportion of positive cases, while the specificity of 0.948 indicates a relatively low false-positive rate.

The classification gain is associated with the fusion of complementary spatial, frequency, and texture information. The CNN branch captured localized morphological patterns, including border irregularity and pigmentation distribution, whereas the transformer branch modelled broader relationships across the lesion region. The wavelet and texture components further contributed information that may be insufficiently represented by a single spatial feature extractor. This result supports the use of multimodal representations for distinguishing visually similar lesion categories.

The proposed model also produced the lowest expected calibration error, with an ECE of 0.021. In comparison, EfficientNet-B7, DenseNet-201, and the Vision Transformer obtained ECE values of 0.067, 0.072, and 0.059, respectively. This result indicates that the confidence scores generated by HEU were more closely aligned with the observed correctness of its predictions. Reliable calibration is especially important in clinical decision support because high classification accuracy alone does not guarantee that probability outputs can be interpreted as meaningful confidence estimates. Figure 5 show the classification performance of the proposed HEU model.

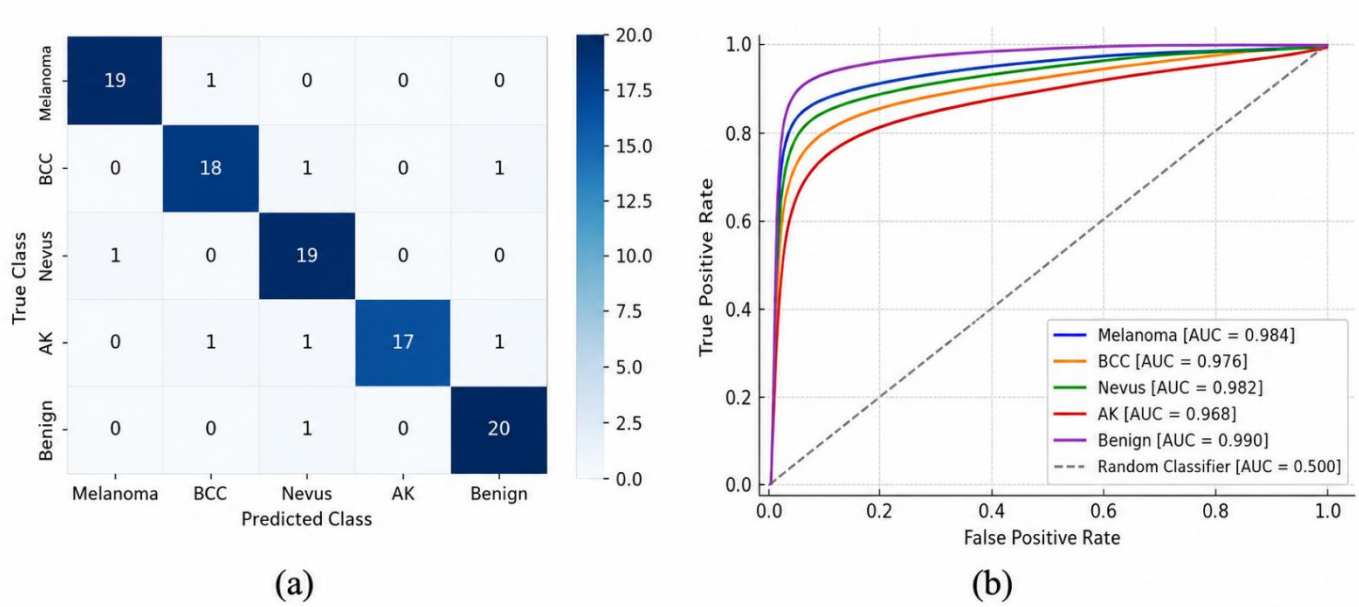

Figure 5. (a) normalized confusion matrix, (b) one-versus-rest ROC curves for each diagnostic class.

Uncertainty analysis showed that 92% of low-confidence predictions corresponded to ambiguous or atypical lesions. This suggests that predictive uncertainty may help identify cases that require human review rather than automatic acceptance. However, this finding should be supported by a clearly defined low-confidence threshold, the number of flagged samples, and an independent assessment of whether the flagged lesions were genuinely ambiguous.

4.3. Explainability, Ablation Analysis, and Overall Findings

The Multi-level Explainability Interface generated visual, feature-level, and textual explanations for each prediction. Grad-CAM was used to identify image regions that contributed most strongly to the predicted class, while SHAP was used to estimate the contribution of individual or grouped features. Representative outputs indicated that the model concentrated on lesion regions containing uneven pigmentation, irregular borders, and asymmetric structures rather than on surrounding background areas (see figure 6).

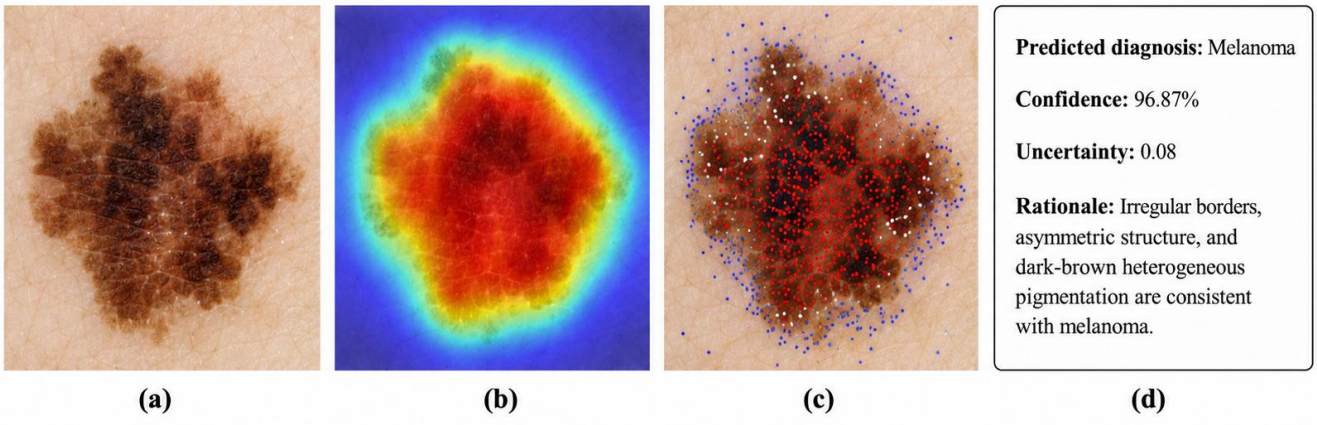

Figure 6. Representative explanation generated by MEI: (a) original lesion image, (b) Grad-CAM attention map, (c) SHAP attribution map, and (d) predicted diagnosis, confidence, uncertainty, and textual rationale.

The alignment between attention maps and the lesion area suggests that the model's decisions were based on diagnostically relevant image regions. However, visual plausibility alone does not establish explanation quality. The final manuscript should quantify explanation localization using lesion masks or expert-annotated regions. Appropriate measures may include intersection over union between saliency regions and lesion masks, pointing-game accuracy, or dermatologist rating scores. The claim that the model reasons similarly to a dermatologist should be avoided unless supported by a formal human evaluation [9], [16].

Overall, the reported results indicate that EDLF-Skin improves both lesion segmentation and disease classification compared with the selected baselines. AMU-Trans achieved the highest overlap-based segmentation scores, while HEU produced the highest classification performance and the lowest calibration error. The combination of lesion-focused preprocessing, multimodal feature representation, complementary classifier branches, and uncertainty estimation provides a coherent explanation for these improvements.

Nevertheless, the current evidence remains incomplete in several areas. The final manuscript must report dataset size and origin, patient-level splitting, repeated-run variability, class-wise results, uncertainty thresholds, and ablation findings. External validation on an independent dataset would further strengthen the claim that the model generalizes across imaging domains, acquisition devices, and patient populations.

4.4. Discussion

The experimental results indicate that EDLF-Skin benefits from integrating lesion segmentation with complementary feature learning. AMU-Trans achieved a Dice coefficient of 0.956 and an IoU of 0.935, outperforming the conventional segmentation models evaluated in this study. This improvement can be attributed to the combination of CNN-based local feature extraction and Transformer-based contextual modelling, which enables the network to capture both fine lesion boundaries and broader spatial relationships. Similar findings have been reported in recent skin-lesion studies, where hybrid architectures combining convolutional and attention mechanisms improved lesion segmentation and classification performance [22], [23]. The subsequent HEU classifier achieved 96.87% accuracy, a 0.953 F1-score, and a 0.981 AUC, suggesting that the integration of spatial, frequency, and texture representations provides complementary information for distinguishing visually similar skin conditions [24].

The uncertainty results provide an additional perspective on the reliability of the proposed model. HEU achieved an ECE of 0.021, which was lower than the three baseline classifiers, indicating that its predicted confidence was better aligned with prediction correctness. This is important in medical diagnosis because a highly accurate model may still produce unreliable decisions when it is excessively confident in incorrect predictions. Recent studies have therefore emphasized the importance of calibrated and uncertainty-aware models for safer clinical decision support [25]. In the present study, 92% of low-confidence predictions were associated with ambiguous or atypical lesions, suggesting that uncertainty estimation could help identify cases requiring additional expert assessment. However, this finding should

 ISSN 2723-6471

be validated using a predefined uncertainty threshold and an independent dataset before being interpreted as a clinically reliable referral mechanism [26].

The explainability results further support the potential clinical usefulness of EDLF-Skin. Grad-CAM highlighted lesion regions associated with irregular pigmentation, asymmetric structures, and border characteristics, while SHAP provided additional feature-level attribution. These findings are consistent with recent research showing that Grad-CAM and SHAP can improve transparency in skin-lesion classification by indicating the regions and features that contribute to model predictions [27]. Nevertheless, visually plausible heatmaps do not necessarily prove that the explanations faithfully represent the model's decision process. Future work should therefore evaluate explanation quality quantitatively using lesion-mask overlap, localization metrics, faithfulness measures, or dermatologist assessment. Overall, EDLF-Skin provides a unified framework that combines diagnostic performance, calibration, uncertainty estimation, and interpretability, although external validation and systematic ablation studies remain necessary before clinical deployment.

5. Conclusion

This study presented EDLF-Skin, an integrated framework for automated skin-disease diagnosis that combines image normalization, lesion segmentation, multimodal feature learning, uncertainty-aware classification, and multilevel explanation. The Physio-Color Consistency module was designed to reduce illumination variation and acquisition artifacts while preserving diagnostically relevant color and texture information. AMU-Trans subsequently localized lesion regions by combining convolutional feature extraction with transformer-based contextual modelling. Spatial, frequency, and texture representations were then integrated through the Multi-Modal Feature Fusion module and classified using the Hybrid Ensemble with Uncertainty.

Based on the reported experiments, AMU-Trans achieved a Dice coefficient of 0.956 and an IoU of 0.935, outperforming U-Net, Attention U-Net, and SegNet. The complete HEU classifier obtained an accuracy of 96.87%, an F1-score of 0.953, an AUC of 0.981, and an expected calibration error of 0.021. These results indicate that the proposed framework not only improved lesion segmentation and diagnostic discrimination but also produced probability estimates that were better aligned with prediction correctness. The explainability interface further supported model interpretation through Grad-CAM visualization, SHAP-based attribution, and concise textual rationales.

The main contribution of this work lies in treating preprocessing, segmentation, representation learning, classification, uncertainty estimation, and explanation as connected stages of a single diagnostic workflow rather than as isolated components. This design enables the model to focus on lesion-specific information, exploit complementary feature domains, and identify predictions that may require expert review. However, the reported findings should be interpreted with several limitations. The final study must clearly document dataset provenance, sample size, patient-level partitioning, repeated-run variability, uncertainty thresholds, ablation results, and the procedure used to evaluate explanation quality.

Future work should validate EDLF-Skin on larger independent datasets collected across different devices, institutions, skin tones, and disease categories. Prospective assessment involving dermatologists is also needed to determine whether the generated explanations and uncertainty estimates improve diagnostic confidence, referral decisions, and clinical workflow. With these validations, the framework may serve as a more reliable foundation for computer-assisted dermatology and teledermatology applications.

6. Declarations

6.1. Author Contributions

Conceptualization: R.K.M., C.B.; Methodology: R.K.M., A.N.S.; Software: C.B., R.K.M.; Validation: A.N.S., R.K.M.; Formal Analysis: R.K.M.; Investigation: C.B., R.K.M.; Resources: A.N.S., C.B.; Data Curation: C.B.; Writing – Original Draft Preparation: R.K.M.; Writing – Review and Editing: A.N.S., C.B.; Visualization: C.B.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. P. Choy, B. J. Kim, A. Paolino, W. R. Tan, S. M. L. Lim, J. Seo, S. P. Tan, L. Francis, T. Tsakok, M. A. Simpson, J. N. W. N. Barker, M. D. Lynch, M. S. Corbett, C. H. Smith, and S. K. Mahil, "Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease," *npj Digit. Med.*, vol. 6, no. 1, pp. 1–12, Aug. 2023, doi: 10.1038/s41746-023-00914-8.

[2] S. Belarouci, M. A. Chikh, A. Belaidi, Z. A. Elaouaber, and S. H. Cherif, "Deep learning for skin disease classification using CNNs and transfer learning," *AICPS*, vol. 2025, no. Mar., pp. 91–96, 2025, doi: 10.1109/AICPS66617.2025.11513493.

[3] S. M. Saqib, M. Iqbal, M. T. B. Othman, T. Shahazad, Y. Y. Ghadi, S. Alamro, and T. Mazhar, "Lumpy skin disease diagnosis in cattle: A deep learning approach optimized with RMSProp and MobileNetV2," *PLoS One*, vol. 19, no. 8, pp. 1–15, Aug. 2024, doi: 10.1371/journal.pone.0302862.

[4] S. Pingulkar, D. Divekar, and A. Tiwary, "Deep learning for skin disease diagnosis with end-to-end data security," *ICI*, vol. 2023, no. Dec., pp. 1–10, 2023, doi: 10.1109/ICI60088.2023.10421188.

[5] A. Kumar, A. K. Sahoo, N. K. Bhaul, S. Nayak, A. Rath, and G. Panda, "Skin disease diagnosis system using machine and deep learning algorithms," *SCOPES*, vol. 2024, no. Dec., pp. 1–10, 2024, doi: 10.1109/SCOPES64467.2024.10990791.

[6] K. Vayadande, O. Lohade, S. Umbare, P. Pise, A. Matre, M. Mule, I. Mahajan, V. Karanjawane, M. Ingale, B. Gaikwad, and V. Karpe, "Automated multiclass skin disease diagnosis using deep learning," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 11, pp. 327–336, 2024.

[7] M. Karnati, P. Sai, R. K. Pittala, K. Radhika, and K. Sugamya, "Deep learning-based predictive skin disease diagnosis and analysis," *ICKECS*, vol. 2024, no. May, pp. 1–10, 2024, doi: 10.1109/ICKECS61492.2024.10617139.

[8] M. Subramanian, K. Shanmugavadivel, S. Thangaraj, J. Cho, and S. VE, "Attention-aware deep learning models for dermoscopic image classification for skin disease diagnosis," *Curr. Med. Imaging*, vol. 21, no. 11, pp. 1–12, 2025, doi: 10.2174/0115734056332443241129113146.

[9] M. Singh, M. Singh, D. De, S. Handa, R. Mahajan, and D. Chatterjee, "Towards diagnosis of autoimmune blistering skin diseases using deep neural network," *Arch. Comput. Methods Eng.*, vol. 30, no. 6, pp. 3529–3557, Aug. 2023, doi: 10.1007/s11831-023-09910-3.

[10] R. Shaik, S. K. Bodhapati, A. Uddandam, L. Krupal, and J. Sengupta, "A deep learning model that diagnosis skin diseases and recommends medication," *PCEMS*, vol. 2022, no. Jun., pp. 7–10, 2022, doi: 10.1109/PCEMS55161.2022.9808065.

[11] E. Goceri, "Diagnosis of skin diseases in the era of deep learning and mobile technology," *Comput. Biol. Med.*, vol. 134, no. 2021, pp. 1–10, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104458.

[12] S. Malliga, G. Sherly Infanta, S. Sindoora, and S. Yogarasi, "Skin disease detection and classification using deep learning algorithms," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 3, pp. 255–260, 2020.

[13] V. P. Kolanchinathan, S. S. Babu, T. R. D. Kumar, M. Mari Perumal, N. V. Abhiram, and Y. S. Madhesh Kanna, "Enhancing medical diagnosis with deep learning: Automated detection of Alzheimer's and skin diseases," *RAEEUCCI*, vol. 2025, no. Jun., pp. 1–10, 2025, doi: 10.1109/RAEEUCCI63961.2025.11048350.

[14] S. Verma, A. Kumar, and M. Kumar, "A review on deep learning techniques for skin disease detection," *ACET*, vol. 2024, no. Dec., pp. 1–10, 2024, doi: 10.1109/ACET61898.2024.10729965.

[15] S. Fatima, M. U. Akram, S. Mohammad, and S. B. Ahmed, "Deep learning in dermatopathology: Applications for skin disease diagnosis and classification," *Discov. Appl. Sci.*, vol. 7, no. 9, pp. 1–15, Sep. 2025, doi: 10.1007/s42452-025-07138-3.

[16] L. Tyagi, D. Singh, and N. Goyal, "The deep learning for skin disease diagnosis and classification: A review of cutting-edge techniques, outcomes, and limitations at a glance," *AIP Conf. Proc.*, vol. 2024, no. Nov., pp. 1–12, 2024, doi: 10.1063/5.0234321.

[17] H. Van Thanh, N. D. Quang, and K.-H. Jo, "Modern deep learning models for skin disease diagnosis," *IWIS*, vol. 2023, no. Oct., pp. 1–10, 2023, doi: 10.1109/IWIS58789.2023.10284645.

[18] A. K. Gupta, V. Baviskar, P. Sethi, and D. Bhise, "Application of deep learning in skin disease diagnosis: A review," *ICICT*, vol. 2023, no. Dec., pp. 1773–1780, 2023, doi: 10.1109/ICICT57646.2023.10134158.

[19] G. Al-Gaphari, N. A. M. Morshed, and K. AlAidarous, "A deep learning-based approach for a real-time diagnosis of skin diseases," *Neural Comput. Appl.*, vol. 37, no. 31, pp. 25945–25972, 2025, doi: 10.1007/s00521-025-11577-9.

[20] P. K. Vasishtamol and P. Soni, "Advances in deep learning for dermatology: Convolutional neural networks for automated skin disease detection and diagnosis," *ICCCNT*, vol. 2023, no. Jul., pp. 1–10, 2023, doi: 10.1109/ICCCNT56998.2023.10306636.

[21] P. Pattnayak, A. Mohanty, S. Patnaik, and D. Mohanty, "A comprehensive study on detecting skin diseases using deep learning techniques," *GIEST*, vol. 2024, no. Dec., pp. 1–10, 2024, doi: 10.1109/GIEST62955.2024.10960070.

[22] M. Fiaz, M. B. Shoaib Khan, A. H. Khan, A. Bilal, M. Abdullah, A. A. Darem, and R. Sarwar, "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification," *Front. Med.*, vol. 12, no. 2025, pp. 1–15, 2025, doi: 10.3389/fmed.2025.1681542.

[23] P. Murali and D. H. Mazumder, "DermaScanAI: An explainable hybrid deep learning framework for automated skin lesion classification using dual attention and metadata fusion," *Sci. Rep.*, vol. 16, no. 2026, pp. 1–15, 2026, doi: 10.1038/s41598-026-46011-0.

[24] S. M. Yasir and H. Kim, "Attention-based deep feature aggregation network for skin lesion classification," *Electronics*, vol. 2025, no. 12, pp. 1–12, 2025, doi: 10.3390/electronics14122364.

[25] M. A. A. Khan, M. Tahir, S. A. Irtaza, M. R. Mughal, and M. N. Yousaf, "SynthraXCoreNet: An interpretable, well-calibrated six-CNN ensemble for dermoscopic skin-lesion classification," *IEEE Access*, vol. 14, no. 2026, pp. 6287–6309, 2026, doi: 10.1109/ACCESS.2025.3641868.

[26] R. Singh, R. Gorantla, S. G. Allada, and P. Narra, "SkiNet: A deep learning framework for skin lesion diagnosis with uncertainty estimation and explainability," *PLoS One*, vol. 17, no. 2022, pp. 1–15, 2022, doi: 10.1371/journal.pone.0276836.

[27] R. Y. G. Paccotacya-Yanque, A. Bissoto, and S. Avila, "Are explanations helpful? A comparative analysis of explainability methods in skin lesion classifiers," *SIPAIM*, vol. 2024, no. Nov., pp. 1–5, 2024, doi: 10.1109/SIPAIM62974.2024.10783606.

[28] Y. Gulzar, B. I. Ya'u, M. Alkanan, and C. W. Onn, "ScNet: A lightweight CNN with depthwise and SE modules for skin lesion classification," *Comput. Methods Biomech. Biomed. Eng.: Imaging Vis.*, vol. 13, no. 1, pp. 1–12, 2025.