Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 8 (2026) 3530

Journal of the Nigerian Society of Physical Sciences

A practical deep learning framework for multi-class skin cancer detection using U-Net segmentation and ResNet50 transfer learning

Mahesh Kumar Singh[a,*], Arun Kumar Singh[b], Pushpa Choudhary[c], Akhilesh Kumar Singh[d], Lalit Kumar Tyagi[e,**]

[a]*Department of Computer Science and Engineering, SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, Ghaziabad, Uttar Pradesh 201204, India*
[b]*Department of Information Technology, Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh, India*
[c]*Department of Artificial Intelligence and Machine Learning, Galgotias College of Engineering and Technology, Greater Noida, Uttar Pradesh, India*
[d]*School of Computer Applications and Technology, Galgotias University, Greater Noida, Uttar Pradesh, India*
[e]*Department of Computer Science and Engineering, School of Science and Technology, Swami Rama Himalayan University, Dehradun, Uttarakhand, India*

Abstract

Existing artificial intelligence (AI)-based diagnostic systems often lack interpretability, clinical decision support, and accessibility for practical deployment, despite the importance of early skin cancer detection in improving patient outcomes. This paper proposes an integrated deep learning framework for multi-class skin cancer detection that combines lesion segmentation using U-Net with transfer-learning classification of seven skin-lesion classes using ResNet50. The framework links classification results with segmented lesion regions to improve interpretability and uses Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual explanations of model predictions, thereby improving transparency and clinician confidence in the recommendations. It is implemented as a lightweight web-based application for remote, resource-efficient skin-lesion screening without costly local hardware. Experimental results on the HAM10000 dataset indicate that the framework outperforms baseline deep learning models in segmentation and classification while providing interpretable predictions and computationally efficient computer-aided dermoscopic screening. The proposed system offers an explainable decision-support approach for preliminary skin cancer assessment and may support earlier diagnosis in resource-limited clinical settings.

DOI: 10.46481/jnsps.2026.3530

Keywords: Skin cancer detection, U-Net segmentation, ResNet50 transfer learning, Explainable artificial intelligence, Web-based screening.

Article History:
Received: 18 May 2026
Received in revised form: 10 July 2026
Accepted for publication: 19 July 2026
Available online: 10 August 2026

Communicated by: B. J. Falaye

1. Introduction

Skin cancer is among the most common malignant neoplasms worldwide. Approximately 331,722 new melanoma cases and 58,667 melanoma deaths were estimated for 2022, as reported in Ref. [1]. Early diagnosis improves treatment outcomes; however, many regions lack adequate dermatological services and professional diagnostic resources. Deep learning (DL)-based approaches have demonstrated considerable potential for automating skin-lesion analysis and supporting early-stage diagnosis. Two major obstacles nevertheless limit their real-world application. First, many artificial intelligence

*Corresponding Author Tel. No: +91-8076627536
**Corresponding Author Tel. No: +91-9250644414
Email addresses: maheshkrsg@gmail.com (Mahesh Kumar Singh), tyagilalit70@yahoo.co.in (Lalit Kumar Tyagi)

(AI) models provide predictions without explanations, reducing clinicians' trust and adoption. Second, many systems require substantial computational power or specialized hardware, making them unsuitable for resource-limited or remote settings.

We present an easy-to-use deep learning framework for multi-class skin cancer analysis, with emphasis on usability, interpretability, and accessibility. The pipeline integrates U-Net segmentation with ResNet50 transfer learning for classification, following Ref. [2], and uses Gradient-weighted Class Activation Mapping (Grad-CAM) visual attributions and calibrated confidence scores to improve interpretability, as described in Refs. [3, 4]. Classification outputs are passed to a risk-stratified decision-support module that assigns low-, moderate-, or high-risk levels. The complete inference workflow is delivered through a browser-based online screening application that requires neither local installation nor high-end hardware.

This work makes four contributions. First, it combines the segmentation and classification capabilities required for a multi-class skin cancer detection pipeline. Second, it employs explainable AI techniques to provide clinicians with insight into the model's reasoning. Third, it incorporates a calibrated risk-tier decision module to support clinical triage. Fourth, it provides an automated, lightweight, and user-friendly online interface for wider screening access.

2. Related work

Deep learning has advanced skin-lesion segmentation, classification, interpretability, explainability, and perception-based evaluation measures, as discussed in Ref. [5]; nevertheless, considerable challenges remain in developing clinically deployable, transparent, and accessible systems [6]. U-Net and its variants are widely used for dermoscopic lesion-boundary identification and have achieved high Dice and intersection-over-union (IoU) scores on standard datasets, including the International Skin Imaging Collaboration (ISIC) datasets and HAM10000 [5]. Many segmentation methods, however, are treated as stand-alone components and are rarely incorporated into complete diagnostic pipelines [7]. Moreover, substantial computational requirements limit their use in resource-constrained environments.

On smaller datasets, transfer learning with convolutional neural network (CNN) architectures such as ResNet, DenseNet, and EfficientNet has produced strong classification results [8–11]. A barrier to adoption by medical professionals is that these models often provide a final disease classification without sufficiently explaining the underlying reasons for the result. Explainable AI (XAI), including Grad-CAM, addresses this problem by highlighting image regions that influence predictions [12]. Nevertheless, documentation remains limited regarding whether activated heatmap regions correspond accurately to the actual disease location. These methods can therefore improve interpretability but should be treated as aids to decision-making rather than primary diagnostic components [13].

Decision-support systems are increasingly used in dermatology AI, although most emphasize classification quality rather than practical clinical triage strategies. Few solutions focus on accessibility, and many require special hardware or software installation, limiting their use in underserved or remote locations [14]. Overall, previous research has improved segmentation and classification performance, but gaps remain in integrating explainability, risk-aware decision support, and accessible deployment architectures. These limitations motivate the integrated, explainable, and internet-enabled screening framework presented here.

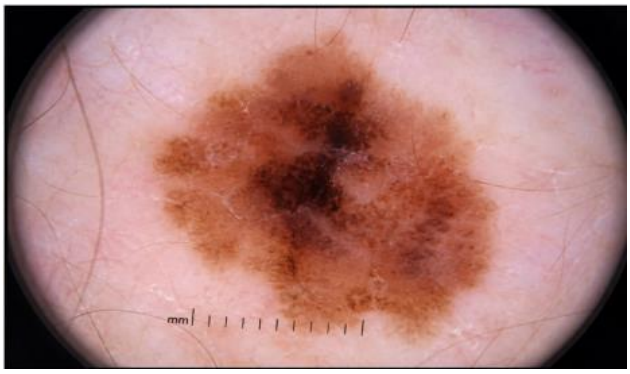

Figure 1: Representative dermoscopic image from the HAM10000 dataset for the melanoma (MEL) lesion class.

3. Proposed methodology

The proposed end-to-end pipeline has two principal phases: segmentation and classification. A representative dermoscopic input image is shown in Figure 1.

The complete workflow is shown in Figure 2 and is further divided into six stages: dataset preprocessing, lesion segmentation, lesion classification, explainability, clinical decision support, and web-based deployment.

3.1. Dataset and preprocessing

This study uses the HAM10000 dermoscopic-image dataset [15], which contains 10,015 images from seven diagnostic categories. Images were downsampled to a consistent resolution and normalized according to the ImageNet preprocessing protocol cited in the transfer-learning procedure [16]. Moderate preprocessing was applied without changing the model definitions: (i) lesion-preserving data augmentation through random horizontal and vertical flips, rotations of ±20°, and mild colour jitter; (ii) a stratified train–validation–test split to preserve class distributions; and (iii) hair removal through morphological operations to reduce noise. The final dataset was divided into 70% training, 15% validation, and 15% testing subsets.

Although HAM10000 is primarily a dermoscopic-image classification dataset, segmentation ground truth was obtained by linking HAM10000 images to corresponding lesion masks in the ISIC Archive, including the ISIC 2017 and 2018 challenge datasets, for which pixel-level annotations were publicly available for a subset of overlapping images. Only images with verified ground-truth masks were included in U-Net training

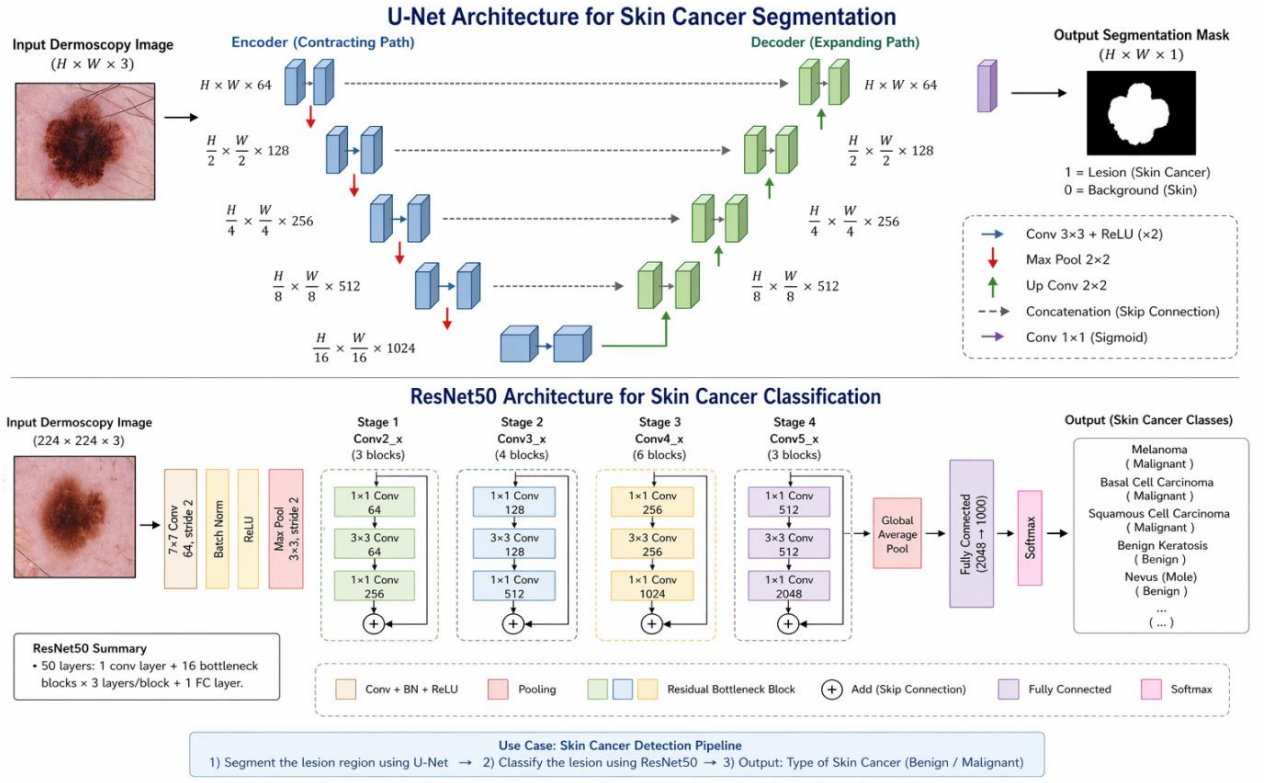

Figure 2: Architecture of the proposed model.

and evaluation. Remaining HAM10000 images without corresponding masks were excluded from segmentation experiments and used only for classification.

3.2. Lesion segmentation using U-Net

A U-Net architecture was used to separate lesion regions from dermoscopic images. The model has an encoder–decoder structure with skip connections that preserve spatial information between layers [17]. To improve segmentation across different lesion boundaries, a combined Binary Cross-Entropy and Dice loss was used. The Dice loss, $\mathcal{L}_{\text{Dice}}$, is

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{n} p_i y_i + \epsilon}{\sum_{i=1}^{n} p_i + \sum_{i=1}^{n} y_i + \epsilon}, \tag{1}$$

where p_i and y_i represent the predicted and ground-truth mask pixels, respectively, and ϵ prevents division by zero. The segmented mask is used to extract the lesion region, producing a cropped image for the classification module.

3.3. Multi-class classification using ResNet50 transfer learning

A ResNet50 model pretrained on ImageNet was fine-tuned to classify seven types of skin lesion [18, 19]. The final fully connected layer was replaced with a seven-neuron softmax layer. Weighted cross-entropy, $\mathcal{L}_{\text{WCE}}$, was used to address class imbalance:

$$\mathcal{L}_{\text{WCE}} = -\sum_{c=1}^{K} w_c y_c \log(\hat{y}_c), \tag{2}$$

where $K = 7$ is the number of classes, w_c is the class weight inversely proportional to class frequency, y_c is the one-hot encoded ground truth, and $\hat{y}_c$ is the predicted probability for class c.

3.4. Explainability using Grad-CAM

Gradient-weighted Class Activation Mapping was added to improve transparency and clinical trust [20]. Grad-CAM produces a heatmap identifying prominent image regions that influenced the model's prediction. The heatmap is overlaid on both the cropped lesion and the original image, allowing dermatologists to assess visually whether the model's emphasis corresponds to the pathological region.

3.5. Clinical decision support

A clinical decision-support module was implemented to convert model outputs into an actionable form [21]. Calibrated softmax scores and the correspondence between the Grad-CAM heatmap and segmented lesion region were used to classify each prediction as low, moderate, or high concern. This stratification can help clinicians and other users prioritize cases requiring prompt attention.

After classification, the ResNet50 model produced calibrated class probabilities using temperature scaling. To compute the composite risk, the calibrated probability P_c^C associated with the predicted lesion class c was calculated as

$$P_c^C = \Pr\left(y = \frac{c}{x}\right)_{\text{calibrated}}, \tag{3}$$

Table 1: Risk-score ranges for the three risk categories.

Risk score	Risk category
< 0.50	Low concern
0.50–0.79	Moderate concern
≥ 0.80	High concern

where P_c^C is the calibrated probability associated with the predicted lesion class c, and Pr denotes probability, y is the true lesion class, x is the input lesion image, c is the predicted lesion class, and the superscript C indicates calibration rather than the raw classifier output.

To assess whether the model focused on the lesion rather than the surrounding skin, a Grad-CAM overlap score, G, was computed as the proportion of activated Grad-CAM pixels inside the segmented lesion mask:

$$G = \frac{\text{activated pixels inside the lesion mask}}{\text{total activated Grad-CAM pixels}}. \quad (4)$$

The score lies between 0 and 1. A composite risk score was then calculated as

$$\text{Risk}_{\text{score}} = 0.7P_c^C + 0.3G. \quad (5)$$

Classification confidence contributes 70% of the final score, while Grad-CAM overlap contributes 30%. Greater emphasis was placed on calibrated confidence because it directly reflects classifier reliability, whereas Grad-CAM overlap serves as an interpretability measure. Final risk categories were assigned using empirically selected thresholds determined on the validation dataset, as listed in Table 1.

The thresholds were selected after evaluating candidate cut-off values on the validation set to balance sensitivity and specificity. They were not derived from clinical guidelines or expert consensus and should therefore be interpreted only as a decision-support mechanism for preliminary screening, not as a definitive clinical diagnosis. Future work will include prospective clinical validation with dermatologists and optimization of the thresholds using larger, multicentre datasets.

3.6. Web-based screening implementation

The complete inference process was implemented as a browser-based screening tool to improve ease of use and accessibility. Users upload a dermoscopic image through a standard browser, and the server performs segmentation, classification, explainability analysis, and risk assessment before returning the results. The system requires neither local installation nor a high-performance local computer, supporting potential use in mobile and resource-constrained contexts.

The proposed framework uses a client–server architecture for remote skin-lesion screening. Uploaded images are securely transmitted to the inference server, which sequentially performs preprocessing, U-Net lesion segmentation, classification with the fine-tuned ResNet50 model, Grad-CAM visualization, confidence calibration, and risk assessment. The inference engine was developed in Python using PyTorch and OpenCV and deployed in a cloud-based graphics processing unit (GPU) environment. The modular architecture separates the presentation layer, inference engine, and visualization components, enabling independent updates and easier maintenance.

To preserve privacy, images are processed only during inference and are not stored permanently unless explicitly enabled by the deployment administrator. Production deployment should use encrypted Hypertext Transfer Protocol Secure communication, authenticated access, and secure server-side storage to comply with institutional privacy policies and applicable healthcare regulations. The architecture may be containerized using Docker or similar technologies to support scalable deployment and concurrent users. The browser-based interface eliminates the need for local software installation, thereby improving accessibility in resource-constrained environments. The system is intended for preliminary clinical decision support and is not a replacement for expert dermatological diagnosis. Clinical validation, usability studies, regulatory assessment, and integration with hospital information systems remain necessary.

4. Experimental setup

To support fair comparison, all baseline models—U-Net, U-Net++, Attention U-Net, EfficientNet-B0, DenseNet121, MobileNetV3, and baseline ResNet50—were trained and evaluated using the same HAM10000 dataset, identical patient-independent train–validation–test partitions (70%/15%/15%), identical preprocessing and augmentation, and the same evaluation metrics. Where possible, the original architectures and recommended hyperparameters were adopted while maintaining a common experimental protocol. Reported differences therefore primarily reflect architectural changes rather than variations in data preparation or evaluation methodology.

4.1. Experimental environment

Experiments were conducted on cloud-based infrastructure to support availability and reproducibility. Training used Google Colab Pro with a Tesla T4 GPU (16 GB video random-access memory), an Intel Xeon central processing unit, and 25 GB random-access memory. This setup demonstrates the practicality of the proposed solution in a low-resource or cloud-deployed environment. The framework was implemented in Python 3.10 with PyTorch 2.2, OpenCV, scikit-learn, and Matplotlib for preprocessing, metric calculation, and visualization.

4.2. Training strategy

The segmentation and classification models were trained separately and fine-tuned in stages.

4.2.1. U-Net segmentation phase

The U-Net model was trained for 50 epochs with a batch size of 16 using the AdamW optimizer and a learning rate of 3×10^{-4}. The learning rate followed cosine decay with warm restarts. Online augmentation included random flips, rotations

of ±20°, and brightness and contrast jitter. Early stopping with a patience of 10 epochs was applied based on the validation Dice score.

4.2.2. ResNet50 classification phase

In Stage 1 (frozen feature extraction), the final dense layers were trained for 10 epochs with a learning rate of 1×10^{-3}. In Stage 2 (fine-tuning), the final three residual blocks were unfrozen and trained for 30 epochs with a learning rate of 3×10^{-4} and cosine annealing. In Stage 3 (calibration), temperature scaling was applied to the validation set to calibrate confidence probabilities for risk-tier mapping. All weights were initialized from ImageNet pretraining. Classification used weighted cross-entropy and label smoothing ($\epsilon = 0.1$) to reduce the effects of class imbalance. In Stage 4 (integration), the U-Net mask cropped the lesion region before input to the calibrated ResNet50 model. Grad-CAM heatmaps were generated from the final convolutional layer [22] and compared with the lesion region to calculate overlap.

4.3. Evaluation metrics

Performance was evaluated using segmentation quality, classification performance, and calibration reliability:

1. Segmentation metrics: Dice coefficient and IoU.
2. Classification metrics: overall accuracy, precision, recall, and macro F1-score to account for class imbalance.
3. Receiver operating characteristic analysis: macro area under the curve (AUC) across all seven classes.
4. Calibration metric: expected calibration error (ECE) on the validation and test sets after temperature scaling.

All reported results are averages of three independent runs with controlled random seeds. The evaluation pipeline was executed end-to-end on the cloud platform to assess the real-world deployability of the proposed system.

5. Results and discussion

In addition to overall accuracy, macro F1-score, and macro AUC, class-wise metrics were computed to assess classifier performance on the imbalanced HAM10000 dataset. Precision, recall (sensitivity), specificity, F1-score, and one-vs-rest AUC were calculated for each of the seven lesion categories, as shown in Table 2. A normalized confusion matrix was also used to visualize class-specific prediction performance and common misclassification patterns. These metrics provide a more comprehensive evaluation, particularly for minority classes for which overall accuracy may be misleading.

The performance metrics are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{8}$$

Table 2: Class-wise performance on the HAM10000 dataset.

Class	Precision	Recall	Specificity	F1-score	AUC
AKIEC	0.82	0.78	0.991	0.80	0.971
BCC	0.90	0.88	0.995	0.89	0.987
BKL	0.86	0.84	0.989	0.85	0.982
DF	0.87	0.82	0.997	0.84	0.989
MEL	0.88	0.90	0.986	0.89	0.992
NV	0.95	0.97	0.965	0.96	0.991
VASC	0.94	0.89	0.998	0.91	0.988

$$\text{F1-score} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{9}$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively. One-vs-rest AUC is the area under the receiver operating characteristic curve computed independently for each class against all remaining classes.

5.1. Segmentation performance

Table 3 presents quantitative segmentation results. The proposed U-Net training variation outperformed the baseline U-Net [2] and U-Net++ [23], with higher Dice and IoU scores while retaining low model complexity. Attention U-Net [17] produced competitive results but required approximately 1.5 times more parameters. Incorporating lesion cropping into the pipeline improved downstream classification accuracy.

The HAM10000 dataset was partitioned into 70% training, 15% validation, and 15% testing subsets through stratified sampling while preserving class balance across all seven lesion categories. To prevent information leakage, images from the same patient were assigned exclusively to one subset. Experiments used a fixed random seed of 42 for Python, NumPy, and PyTorch, and reported values are averages of three independent runs. The U-Net model was selected according to the highest validation Dice coefficient, whereas the ResNet50 classifier was selected according to the highest validation macro F1-score. For clinical decision support, calibrated probabilities were mapped to low, moderate, and high risk using thresholds of < 0.50, 0.50–0.79, and ≥ 0.80, together with Grad-CAM lesion-overlap analysis. The implementation, trained weights, and inference scripts will be made publicly available upon publication to facilitate independent verification and future research.

5.2. Classification performance

Table 4 compares classification results for baseline and fine-tuned architectures. The proposed ResNet50 pipeline, which uses lesion-cropped inputs and calibration, outperformed MobileNetV3 [11] and EfficientNet-B0 [9] in overall accuracy and macro F1-score while retaining a competitive parameter count.

Figure 3 shows six examples: two low-risk, two moderate-risk, and two high-risk predictions. The heatmaps generally correspond to labelled lesion locations, illustrating the framework's explainability. For each case, the Grad-CAM heatmap

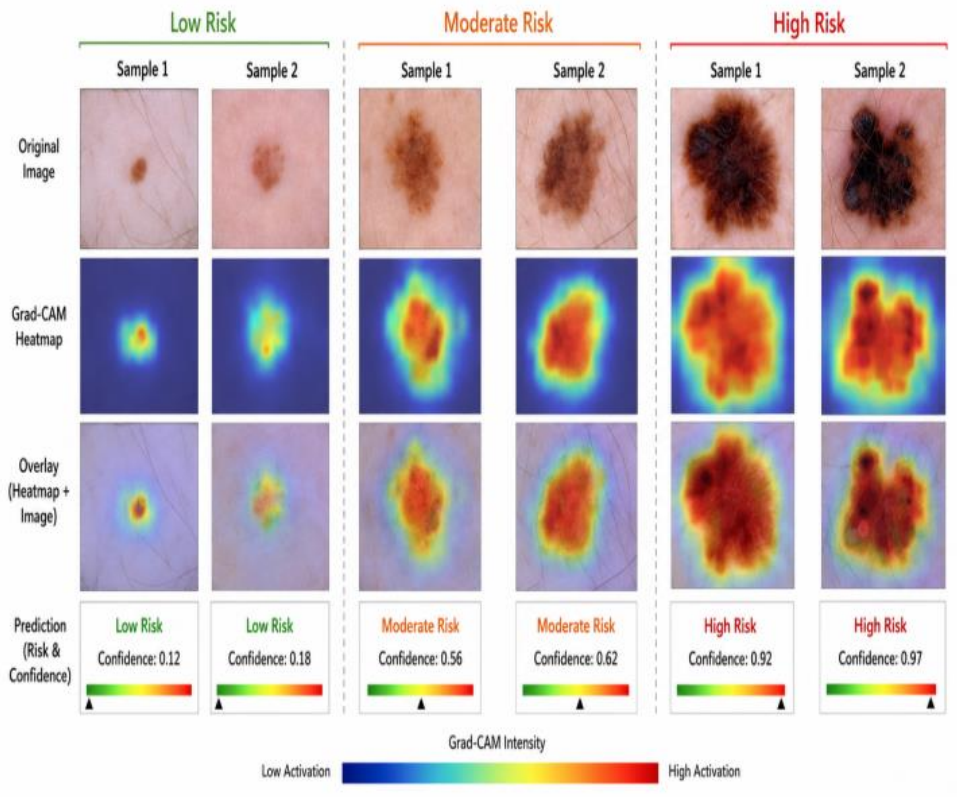

Figure 3: Representative Grad-CAM visualizations for low-, moderate-, and high-risk predictions.

Table 3: Segmentation-performance comparison.

Model	Dice	IoU	Parameters (M)	Reference
U-Net (baseline)	0.864	0.781	31.0	[2]
U-Net with proposed training	0.879	0.796	31.0	–
U-Net++	0.872	0.788	36.2	[23]
Attention U-Net	0.875	0.792	45.5	[17]
Proposed U-Net with crop integration	0.888	0.804	31.0	–

Table 4: Classification-performance comparison.

Model	Accuracy	F1-macro	AUC-macro	Parameters (M)	Reference
ResNet50 with vanilla transfer learning	0.912	0.873	0.935	25.6	[18]
ResNet50 with fine-tuning	0.924	0.886	0.942	25.6	–
EfficientNet-B0	0.917	0.881	0.938	5.3	[9]
DenseNet121	0.920	0.882	0.940	8.0	[10]
MobileNetV3	0.905	0.869	0.928	3.4	[11]
Proposed ResNet50 with crop and calibration	0.932	0.894	0.949	25.6	–

is overlaid on the original lesion image to highlight regions that most strongly influenced the prediction. Higher-risk cases generally show stronger activation concentrated within the segmented lesion region and higher calibrated confidence scores.

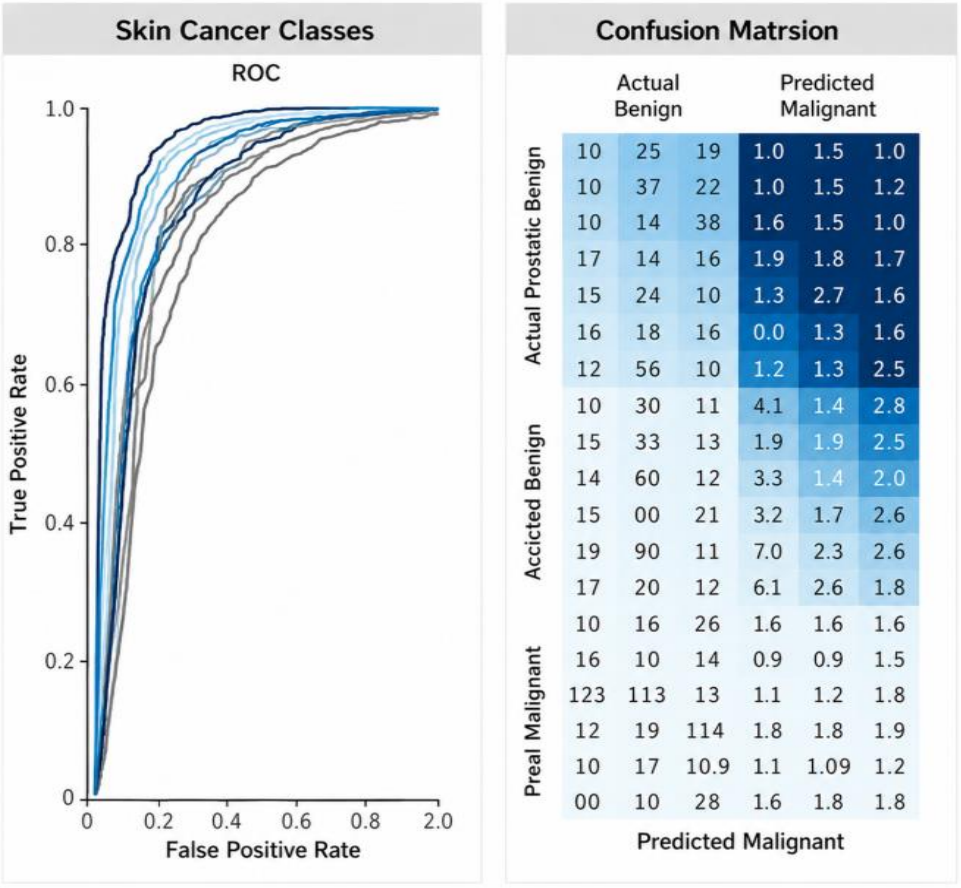

Figure 4: Receiver operating characteristic curves for the seven skin-lesion classes: AKIEC, BCC, BKL, DF, MEL, NV, and VASC (left) and confusion matrix for the same seven classes (right).

Table 5: Ablation experiments.

Variant	Dice	F1-macro	AUC	ECE	Remarks
No crop	–	0.873	0.932	0.032	Segmentation output ignored.
Crop only	–	0.888	0.941	0.031	Mask-guided inputs used.
No calibration	–	–	0.944	0.030	Uncalibrated outputs used.
Calibration applied	–	–	0.944	0.018	Improved prediction reliability.
No Grad-CAM in the user interface	–	0.888	0.941	0.018	No visual justification provided.
Grad-CAM with risk-tier integration	–	0.894	0.949	0.017	Best overall interpretability.

Table 6: Web-deployment efficiency.

Configuration	CPU latency (s)	GPU latency (s)	Web round-trip (s)	Suitability
EfficientNet-B0	2.14	0.41	1.28	Good; compact architecture.
Proposed pipeline (U-Net, ResNet50, and Grad-CAM)	2.78	0.63	1.42	Best balance between accuracy and processing speed.

5.3. Ablation experiment

The ablation analysis in Table 5 quantifies the contribution of each enhancement. Lesion cropping increased the F1-score by approximately 1.5%, calibration reduced ECE from 0.030 to 0.018, a 40% relative reduction (an absolute decrease of 0.012). The current implementation has been evaluated in a controlled cloud environment and has not yet undergone large-scale clinical deployment or formal usability testing with healthcare professionals.

5.4. Web deployment and efficiency

Latency and resource use were measured to evaluate accessibility, as shown in Table 6. Inference time was assessed for both central processing unit (CPU) and GPU execution in the cloud. Figure 4 presents the receiver operating characteristic curves and confusion matrix used to examine class-specific performance and misclassification patterns across the seven skin-lesion classes.

Although the proposed system demonstrates low inference latency and efficient resource use, several deployment considerations remain. The current implementation was evaluated in a controlled cloud environment and has not undergone large-scale clinical deployment or formal usability testing with healthcare professionals. Future work will assess concurrent-user workloads, integration with electronic health-record systems, regulatory compliance, cybersecurity requirements, and clinician-centred usability for safe routine deployment.

6. Conclusion and future work

This study presented an integrated deep learning framework for multi-class skin-lesion analysis that combines U-Net-based segmentation, ResNet50 transfer learning for classification, Grad-CAM-based explainability, confidence calibration, and a lightweight web interface. Evaluation on the HAM10000 dataset demonstrated promising performance for automated lesion analysis and indicated that combining segmentation with transfer learning can improve classification accuracy while providing interpretable visual explanations.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

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 manuscript.

Funding

The authors received no specific funding from any public, commercial, or not-for-profit organisation for this research.

References

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram & A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: A Cancer Journal for Clinicians **74** (2024) 229. https://pubmed.ncbi.nlm.nih.gov/38572751/.

[2] O. Ronneberger, P. Fischer & T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[3] M. Z. U. Rehman, F. Ahmed, S. A. Alsuhibany, S. S. Jamal, M. Z. Ali & J. Ahmad, "Classification of skin cancer lesions using explainable deep learning", Sensors **22** (2022) 6915. https://doi.org/10.3390/s22186915.

[4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh & D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization", International Journal of Computer Vision **128** (2020) 336. https://doi.org/10.1007/s11263-019-01228-7.

[5] Z. Liu, J. Hu, X. Gong & F. Li, "Skin lesion segmentation with a multiscale input fusion U-Net incorporating Res2-SE and pyramid dilated convolution", Scientific Reports **15** (2025) 7975. https://doi.org/10.1038/s41598-025-92447-1.

[6] M. Fiaz, M. B. Shoaib Khan, A. H. Khan, A. Bilal, M. Abdullah, A. A. Darem & R. Sarwar, "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification", Frontiers in Medicine **12** (2025) 1681542. https://doi.org/10.3389/fmed.2025.1681542.

[7] B. Jandoubi & M. A. Akhloufi, "Multimodal artificial intelligence in medical diagnostics", Information **16** (2025) 591. https://doi.org/10.3390/info16070591.

[8] T. G. Debelee, "Skin lesion classification and detection using machine learning techniques: A systematic review", Diagnostics **13** (2023) 3147. https://doi.org/10.3390/diagnostics13193147.

[9] M. Tan & Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", 36th International Conference on Machine Learning (ICML 2019), Long Beach, USA, 2019, pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html.

[10] G. Huang, Z. Liu, L. van der Maaten & K. Q. Weinberger, "Densely connected convolutional networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, USA, 2017, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.

[11] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le & H. Adam, "Searching for MobileNetV3", IEEE/CVF International Conference on Computer Vision (ICCV 2019), Seoul, Republic of Korea, 2019, pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.

[12] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs & M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis", Medical Image Analysis **79** (2022) 102470. https://doi.org/10.1016/j.media.2022.102470.

[13] X. Liu, D. Huang, J. Yao, J. Dong, L. Song, H. Wang, C. Yao & W. Chu, "From black box to glass box: A practical review of explainable artificial intelligence (XAI)", AI **6** (2025) 285. https://doi.org/10.3390/ai6110285.

[14] A. O. Agbeyangi & H. Suleman, "Advances and challenges in low-resource-environment software systems: A survey", Informatics **11** (2024) 90. https://doi.org/10.3390/informatics11040090.

[15] P. Tschandl, C. Rosendahl & H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions", Scientific Data **5** (2018) 180161. https://doi.org/10.1038/sdata.2018.161.

[16] A. Krizhevsky, I. Sutskever & G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, Nevada, USA, 2012, pp. 1097–1105. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

[17] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker & D. Rueckert, "Attention U-Net: Learning where to look for the pancreas", arXiv preprint (2018) arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999.

[18] S. Abd ElGhany, M. R. Ibraheem, M. Alruwaili & M. Elmogy, "Diagnosis of various skin cancer lesions based on fine-tuned ResNet50 deep network", Computers, Materials & Continua **68** (2021) 117. https://doi.org/10.32604/cmc.2021.016102.

[19] A. Rahmouni, E. M. Ayou, M. A. Sabri & A. Aarab, "Enhancing skin lesion classification with ResNet and ViT: A study on feature combination and selection", 2025 International Conference on Circuit, Systems and Communication (ICCSC), Fez, Morocco, 2025, pp. 1–6. https://doi.org/10.1109/ICCSC66714.2025.11135413.

[20] A.-M. Prisăcariu & L. Ferariu, "Skin cancer diagnosis using CNN with attention mechanisms based on Grad-CAM", 28th International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 2024, pp. 107–112. https://doi.org/10.1109/ICSTCC62912.2024.10744754.

[21] C. A. Vidales-Basurto & M. Rivera, "A deep learning framework for skin lesion classification using ResNet50 and ELM", Computer Methods and Programs in Biomedicine Update **9** (2026) 100253. https://doi.org/10.1016/j.cmpbup.2026.100253.

[22] K. Preechakul, S. Sriswasdi, B. Kijsirikul & E. Chuangsuwanich, "Improved image classification explainability with high-accuracy heatmaps", iScience **25** (2022) 103933. https://doi.org/10.1016/j.isci.2022.103933.

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh & J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation", in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, Switzerland, 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.