



OPEN Advancing skin cancer diagnosis with deep learning and attention mechanisms

Yuan Yao¹✉, Umesh Kumar Lilhore²✉, Sarita Simaiya^{2,7}, Sultan M. Aldossary³, Lidia Goso Tekeste⁴✉, Ehab Ghith⁵ & Hanaa A. Abdallah⁶

Skin cancer, particularly melanoma, remains one of the most lethal diseases globally due to challenges in early detection and diagnosis. Conventional image segmentation models often face difficulties due to the high variability in lesion appearance and their limited ability to focus on critical features, thereby compromising diagnostic accuracy. In this study, we introduce an advanced AI-driven framework that integrates a Scaled Dot Attention Mechanism (SDAM) with a modified UNet architecture to improve skin lesion detection. The SDAM, applied as an attention mechanism between the encoder and decoder stages of the UNet, allows the model to prioritize relevant lesion areas and extract essential features while reducing noise. We evaluate the proposed model using the HAM10000 dataset, a diverse collection of skin lesion images, and test it on two additional datasets: ISIC (Preliminary) and PH2 (Preliminary), to assess generalization across various skin lesion types. Our model achieves significant improvements in melanoma detection with Dice scores between 0.97 and 0.988, accuracy ranging from 97.8% to 98.3%, and substantial enhancements in sensitivity. These results outperform baseline models, including standard UNet (Dice score: 0.85, accuracy: 88.4%) and DenseNet (Dice score: 0.87, accuracy: 90.1%). Furthermore, the model's performance was compared to state-of-the-art methods such as Attention UNet, UNet++, and TransUNet, consistently demonstrating superior results. Statistical analysis via a paired t-test reveals a significant performance boost ($p\text{-value} = 0.02$), further validating the effectiveness of the SDAM-enhanced approach. These findings highlight the potential of AI in advancing early skin cancer detection and diagnosis, with the SDAM-UNet framework offering prospects for personalized care and real-time clinical integration. Additionally, our model's performance across multiple metrics such as precision, recall, F1-score, and IoU showcases its robustness in classifying both melanoma and benign skin lesions, reinforcing its utility in clinical practice.

Keywords Melanoma detection, Skin cancer, Image segmentation, AI-powered detection, Attention mechanisms, UNet architecture, Early diagnosis, Lesion classification

Skin Cancer, especially Melanoma, is still a significant public health issue worldwide. Melanoma is one of the fastest-increasing cancers, and there are an estimated 132,000 new cases worldwide each year according to the World Health Organization (WHO)^{1,2}. Early detection of melanoma greatly increases survival rates for patients, but diagnosis is difficult due to a high intra-class variation in lesion appearance, overlap in visual symptoms with benign lesions, and limited consistency in diagnostics between clinical conditions^{3,4}.

Advances in artificial intelligence (AI), especially deep learning, have had promising outcomes in medical image analysis. Such approaches enable automated identification and categorization of lesions with less dependence on subjective human interpretation⁵. Nevertheless, such approaches struggle to generalize to a broader range of lesion types, lack interpretation, and fail to discriminate diagnostically relevant features in an

¹UniSQ Springfield Education City, 37 Sinnathamby Blvd, Springfield Central, QLD 4300, Australia. ²School of Computing Science and Engineering, Galgotias University, Greater Noida, UP, India. ³Department of Computer Engineering and Information, College of Engineering in Wadi AlDawasir, Prince Sattam University, Al-Kharj, Saudi Arabia. ⁴Research Scholar at Eritrea Institute of Technology, Mai-Nefhi College, Himbirti, Mai Nefhi, Eritrea.

⁵Department of Mechatronics, Faculty of Engineering, Ain Shams University, Cairo 11566, Egypt. ⁶Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. ⁷ School of Computer Applications & Technology, Galgotias University, Greater Noida, UP, India. ✉email: yuan.yao@my.jcu.edu.au; umeshlilhore@gmail.com; lidiagosyekeste@gmail.com

image^{6,7}. Furthermore, the physiological context or temporal information integration has been rarely considered in many existing models, and they could help to enhance diagnostic reliability⁸.

Motivation for the study

The increasing load of melanoma worldwide, as well as the drawbacks of traditional diagnostic methods, emphasize the necessity of novel and improved diagnostic tools enabling more robust, sensitive, and specific tumor detection^{7,9}. Expertise in dermatology is not always accessible, particularly not in rural or resource-poor environments, while even experienced experts cannot always decide with great confidence whether a lesion is benign or malignant¹⁰.

Motivated by the need for deep learning models capable of providing high accuracy and interpretability, we conduct this study¹¹. The conventional UNet-type segmentation models cannot give concentrated attention to regions specific to the lesion, which is required for finding early abnormalities leading to pathology. To mitigate these issues, in this study, we design a Scaled Dot Attention Mechanism (SDAM) to fill the missing gap by guiding the model's attention on diagnostically informative regions, which can contribute towards more discriminative feature extraction and less background noise^{11,13}. In addition, the lack of generalization in DNNs trained with single-source data makes multi-dataset evaluations essential. Our method verifies that the implementation can generalize to real clinic scenarios by conducting real clinical experiments on three different datasets (shot and image size variation among them): HAM10000, ISIC (Preliminary), and PH2 (Preliminary)^{14,15}.

Challenges in current methods

Existing segmentation models, like the traditional UNet and DenseNet structures, usually uniformly respond to all pixels in an image (this dilutes focus on clinically important aspects of the photographs, such as lesion borders, asymmetry, and color irregularity as well)^{13,16}. These models also do not have the capabilities to emphasize or emphasize and concentrate on complex lesion morphologies, which are critical to diagnosis. In addition, most methods are trained and tested on single-source datasets, which would result in poor generalization performance to other populations and devices¹⁷.

Class imbalance and the scarcity of some lesion types compound the fragile model generalization, which also degrades performance detrimentally in practical scenarios. The most widely utilized HAM10000 dataset is complete but biased due to overrepresentation of lighter skin tones, potentially leading to a diagnostic performance gap among some demographic groups. These biases raise significant ethical questions and affect clinical utility^{18,19}.

Proposed solution: SDAM-UNet

Some limitations, and we propose a novel deep learning network to incorporate the Scaled Dot Attention Mechanism (SDAM) with a modified UNet. SDAM is presented between the encoder phase and decoded phrases, which encourages SD to highlight diagnostically useful lesion regions and suppress irrelevant background noise. The attention-enriched architecture greatly enhances feature-extracting ability, thus increasing the accuracy of segmentation and reliability for classification^{11,20}.

We tested our model on the HAM10000²⁹ dataset and then further validated it by two independent datasets, i.e., ISIC (Preliminary)³⁰ and PH2 (Preliminary)³¹, to test whether it has generalizability over different lesion types and imaging situations. This multi-dataset assessment methodology provides wider applicability to practical clinical usage.

Novel contributions

The main contributions of this work are:

- *Integration of SDAM into UNet* We added scalable attention to balance feature focus in the UNet model to obtain more accurate segmentations and classifications.
- *Multi-dataset evaluation* To validate whether the models generalize and are stable across HAM10000, ISIC, and PH2 datasets or not.
- *Comprehensive performance metrics* Reporting extensive performance metrics such as Dice score, accuracy, precision, sensitivity, specificity, F1-score, and IoU.
- *Comparison with state-of-the-art models* Superior performance against baseline methods such as UNet, DenseNet, and advanced architectures such as Attention UNet, U-Net++, and TransUNet.
- *Dataset bias* Adapting for data imbalances and dataset bias by preprocessing, augmenting, or balanced sampling.
- *Statistical justification* Use of paired t-test and k-fold cross-validation to statistically justify the robustness and significance of the model proposed.

Ethical considerations

We are aware of the ethical concerns surrounding dataset bias, specifically a lack of representation from people with diverse skin tones in publicly available datasets like HAM10000. These imbalances can lead to the performance of the model degrading for minority populations. Our future work will include training on more balanced datasets and the use of fairness-aware learning methods to achieve equitable diagnostic performance in all skin tones and demographics.

Paper roadmap

The remainder of this paper is organized as follows:

- Related work on skin lesion classification and seg- segmentation is introduced in Sect. 2.
- In Sect. 3, we present the proposed methodology, which consists of data preprocessing, augmentation, network architecture, and training strategy.
- Section 4 describes the experimental results: performance measure, statistical analysis, and visual evaluation.
- Section 5 provides a summary and future directions for further development of AI-assisted dermatological diagnosis methods.

Related work

This section reviews existing deep-learning methods in skin cancer analysis.

Skin lesion detection using deep learning models

Ali¹ presented a web application for monkeypox (mpox) skin lesion detection using state-of-the-art deep learning models. This approach ensures consideration of racial diversity, a significant limitation in most machine learning solutions, as they are often not designed to work well for people with darker skin. The authors successfully adapt to different skin tones by utilizing advanced CNNs and optimization approaches, which employ lesion detection to enhance accuracy. We evaluate the system on a custom dataset that covers a wide range of skin tones and demonstrate that it predicts with both accuracy and equity. However, there are still challenges in ensuring the system's reliability across a broader range of skin conditions and lesions. Zhang⁸ introduced a novel skin cancer detection method that fuses GRU networks and the Orca predation optimization algorithm. The GRUs are very effective in processing sequential medical data, and the Orca algorithm further increases the number of model parameters to improve performance. This paper presents a hybrid technique for developing a novel and efficient model for skin cancer detection. Results indicate high classification accuracy and speed on publicly available datasets of skin cancer images (e.g., ISIC) compared to traditional CNN-based methods.

Naeem⁹ presented the SNC_Net, a hybrid handcrafted and deep learning-based feature model for skin cancer detection. This strategy combines traditional image processing techniques, such as texture and color analysis, with the application of deep learning techniques, particularly convolutional neural networks (CNNs). This combined feature type assures better classification accuracy, particularly when deep learning may not prove effective. The proposed SNC_Net outperforms traditional deep learning methods on benchmark dermoscopic image datasets in complex pattern recognition and fine discrimination of benign and malignant endogenous lesions. Kandhro¹⁰ evaluated the performance of an improved version of the VGG19 model, called E-VGG19, in the real-time diagnosis of skin cancer. E-VGG19 is extended to provide robustness and optimize its performance for real-time applications such as dropout layers and attention modules. In comparison with the skin cancer data from ISIC, this model achieves better performance and faster speed, making it also suitable for real-time applications. Despite this potential, the E-VGG19 model can still overfit when only small datasets are used.

Wu¹¹ also presented MHorUNet, a modified version of UNet, optimized for skin lesion segmentation. The key advantage of this model is the incorporation of high-order spatial interactions that enable a deeper model understanding of complex context relationships in the lesion region. MHorUNet improves segmentation accuracy with the attention mechanism and refines the spatial characteristics of the outstanding border of lesions, such as irregular and ambiguous lines. It achieves a significant improvement over conventional UNet models when tested on ISIC data. Lilhore³⁷ presented SkinEHDLF, a hybrid deep learning model for better diagnosing skin cancer in complex systems. The authors blend classical deep learning and advanced feature extraction techniques to gain high classification accuracy, especially in complex skin lesion types. Their system, tested across a range of dermoscopic images, focuses on being robust to the varying appearances of lesions in order to improve confidence in diagnosis. These results establish the model's practical use in clinical settings, where precise skin cancer diagnosis is still difficult due to varied lesion types and presentations across populations.

Advanced skin lesion classification and segmentation techniques

Sulthana¹³ introduced a deep ensemble of a fully convolutional neural network for skin lesion classification. The proposed model is entirely automatic, requiring minimal expert intervention. It utilizes multiple convolutional and pooling layers, followed by fully connected layers for lesion classification. The model, trained on the ISIC skin lesion dataset, is particularly effective at distinguishing between malignant and benign lesions. Despite the strong results, the model struggles to distinguish highly similar lesions, where subtle differences are difficult to perceive. Khan¹⁴ proposed a novel approach to identifying and categorizing skin lesions using a fusion-assisted technique. This approach integrates deep learning algorithms, such as CNNs, with machine learning methods, such as SVMs. The fusion scheme combines the local features from SVM models with the global features from CNNs to enhance lesion detection and classification accuracy. The model shows excellent discriminating power and generalization ability in localization and identification tasks using multi-skin lesion datasets, including ISIC. However, it is not ideal for real-time applications requiring high speeds and efficiency.

In their study, Shafiq¹⁴ introduced ViT-GradCAM, a fusion between ViT and Grad-CAM for skin lesion classification. By combining the ViT, known for its ability to capture long-range dependencies of image data, with Grad-CAM, the approach highlights the most discriminative regions of an image for classification. This combination allows for accurate classification and interpretability, which is crucial in medical imaging, where model interpretability is essential. ViT-GradCAM was tested on the ISIC 2020 dataset, and the results demonstrate that it is more accurate and interpretable than typical CNN models. Lilhore³⁸ introduced a novel hybrid model of U-Net and a modified MobileNet-V3 model for skin cancer diagnosis. This model uses hyperparameter optimization for performance improvement, which enhances precision in classification. The work emphasizes the advantages of combining the U-Net's power for segmentation with MobileNet-V3's efficient classification, enabling deployment in clinical settings with limited computational resources. Their study enhances both accuracy and computational speed, crucial for real-time and large-scale clinical applications.

Al-Waisy³⁹ developed an end-to-end deep learning solution for the early detection and classification of skin cancer lesions in dermoscopic images. This method leverages state-of-the-art CNN architectures for segmentation and classification tasks, with high discrimination between malignant and benign lesions. The use of multi-scale feature extraction methods enhances lesion localization, particularly for low-contrast and ill-defined lesions (e.g., early-stage melanoma). Al-Waisy et al.³⁹ introduced a more robust model using an advanced segmentation method and proved that their model surpasses state-of-the-art models in accuracy, positioning it as a promising tool for early skin cancer detection.

Nurse-led models and advanced deep learning frameworks for skin cancer detection

Recent works emphasize the importance of combining healthcare professionals' expertise with deep learning models for improved skin cancer detection. Kattach² systematically reviewed nurse-led models for skin cancer detection, highlighting the crucial role of healthcare workers, especially nursing staff, in improving early detection and patient management. While these models can be effective, there is a need for more work on embedding AI systems to improve diagnostic accuracy in low-resource environments. Ozdemir and Pacal³ proposed a deep learning framework for skin cancer detection by integrating ConvNeXtV2 with focal self-attention modules. Their study demonstrates that merging ConvNeXtV2 with attention mechanisms helps focus on the relevant skin lesion regions. This framework achieves superior classification accuracy and speed compared to conventional methods, particularly when using publicly available datasets such as ISIC.

Nawaz et al.⁴ introduced SNC_Net, a model based on deep learning and conventional image processing methods for skin cancer diagnosis. This hybrid approach, combining hand-engineered features with CNNs, improves model performance in classifying complicated skin lesions, especially in cases where deep learning alone may fail. The authors show that SNC_Net surpasses traditional CNN models on well-known dermoscopic benchmark datasets, offering a more stable diagnosis for complex lesion patterns. Uthayakumar⁵ proposed an RNN-optimized CAD system for skin cancer diagnosis. The model exploits a mixed-order relation-aware design that enhances its ability to capture temporal and spatial relations in skin lesion data. While this technique performs better in recognizing complex patterns for benign and malignant lesions, it may overfit when learning from small datasets. Future work aims to improve the generalization of the model.

Pacal et al.⁶ introduced a hybrid deep learning model combining CNNs and ViTs for early skin cancer detection. The CNN-ViT model significantly outperforms conventional CNN models in capturing long-range dependencies and complex lesion appearances. The study indicates that combining CNNs and ViTs can develop a richer understanding of skin lesions, improving diagnostic accuracy, particularly for melanoma detection. Kaur et al.⁷ explored deeper deep learning architectures for melanoma discrimination, emphasizing advanced neural network structures for skin cancer detection. This study highlights the importance of model interpretability, essential for medical applications. By leveraging state-of-the-art deep learning methods, Kaur et al. show improved discrimination between melanoma and benign pigmented lesions. Table 1 presents a comparative analysis of selected articles for skin cancer detection.

Materials and methods

This section presents the key materials and methods used for skin cancer research.

Proposed model

The proposed model consists of an SDAM with an improved UNet architecture for achieving accurate skin lesion segmentation demonstrated in Fig. 1. Skin cancer, in particular malignant melanoma, remains one of the most common causes of morbidity and mortality around the world. Early detection and accurate diagnosis represent key hurdles to fighting the disease^{14,15}. Despite recent advancements in medical image analysis and machine learning, existing models for skin lesion diagnosis (such as conventional image segmentation algorithms) to analyze the shape, color, or texture of skin lesions face two major challenges due to the high variation of characteristics (e.g., size, shape, and texture) of the lesions. Furthermore, these classical procedures may not well emphasize the most significant attributes in the lesions, which may lead to suboptimal segmentation and diagnostic performance^{16,17}.

In order to solve the above problems, we propose a slightly new attention mechanism, denoted SDAM, under the encoder-decoder architecture of the UNet model. The proposed model can concentrate more on the most essential part by allocating dynamic attention to different regions of the images based on their importance as well, which benefits from SDAM¹⁸. This adaptive attention mechanism enables the model to effectively encode subtle but important information of lesions that are easy to be missed by conventional segmentation methods. Relying on these key regions, our model not only gains an accuracy improvement in the segmentation but also provides more trustworthy information for follow-up diagnostic tasks (e.g., categorizing malignancy versus non-malignancy) than overall segmentation quality^{19,20}.

Furthermore, the improved UNet (UNet++) architecture, which has been demonstrated to perform very well for medical image segmentation, is also improved by the addition of SDAM, specifically in precise lesion location^{21,22}. This is particularly useful for melanoma, a skin cancer that is characteristically defined by an irregular border and heterogeneous surface. The identification and verification of these complicated features is especially important for early detection, since early detection can result in the identification of melanoma at a more treatable stage. With the incorporation of these advanced tools, our model attempts to address the limitations of the existing segmentation methods and to provide a more reliable framework for skin lesion detection, leading to more accurate and timely diagnoses of skin cancer^{23,24}.

Working of the proposed model

The complete working of the proposed models is as follows.

Reference	Key Method	Dataset(s)	Key Innovation	Methodological Gaps	Future Directions	Clinical Applicability	Model Efficiency	Computational Cost	Robustness	Real-time Performance
Ozdemir & Pacal ³	ConvNeXtV2 with Self-attention	ISIC	Integration of ConvNeXtV2 with self-attention for improved melanoma detection	Limited to melanoma detection, lacks multi-class capability	Test on multi-class skin lesion datasets and enable real-time clinical usage	Primarily for melanoma detection	Moderate	Moderate	Moderate	High
Nawaz et al. ⁴	CNN-based Deep Learning	ISIC	Dermoscopic image analysis for skin cancer detection using CNN models	Restricted to dermoscopic images, lacks rare conditions detection	Expand to include rare skin lesions and non-dermoscopic images	Limited to static analysis	Moderate	Moderate	High	Moderate
Pacal et al. ⁵	CNN-ViT Hybrid Model	ISIC	Hybrid approach combining CNNs and ViTs for early skin cancer diagnosis	Focus primarily on melanoma, limited to the ISIC dataset	Evaluate on multi-class skin lesions and extend to clinical settings	Real-world clinical use is feasible	High	High	High	High
Kandhro et al. ¹⁰	E-VGG19	ISIC	Enhanced VGG19 for real-time skin cancer detection	Trade-off between real-time performance and segmentation accuracy	Focus on minimizing latency for mobile and resource-constrained devices	Real-time application possible	High	High	High	High
Li et al. ¹⁶	DSEUNet	ISIC	Lightweight UNet for dynamic space grouping enhancement for skin lesion segmentation	High computational cost for real-time use	Focus on reducing computational overhead and enabling mobile device usage	High efficiency for small devices	Low	Low	High	High
Ahamed et al. ¹⁷	UNet with Attention Guidance	ISIC	Use of attention-guided UNet to enhance segmentation quality	Not evaluated on skin tone diversity, and limited real-world validation	Improve robustness for clinical applications and real-time deployment	Applicable to clinical use	Moderate	Moderate	Moderate	Moderate
Wu et al. ¹¹	MHorUNet	ISIC	High-order spatial interaction model for skin lesion segmentation	Limited to the ISIC dataset, lacks real-world deployment validation	Integration with multimodal datasets and real-time IoT-based sensor data	Limited to image-based analysis	Moderate	Moderate	Moderate	Moderate
Wu et al. ¹²	HSH-UNet	ISIC	Hybrid high-order interactive model for enhanced segmentation performance	Does not generalize well to diverse datasets and real-time applications	Test on diverse datasets and deploy for real-time clinical use	Primarily for research use	High	High	High	Moderate
Al-Waisy et al. ³⁹	Deep Learning Framework	Dermoscopy	Early diagnosis and classification of skin cancer lesions	Only tested on dermoscopic, lacks real-time monitoring	Expand to real-time monitoring and multi-modal sensor integration	Suitable for research and clinical testing	Moderate	High	High	High

Table 1. Comparative analysis of skin cancer detection research with segmentation and classification focus.

Enhanced U-net architecture

U-Net architecture with the encoder part of the contracting path (the feature extraction path in U-Net is referred to as the encoder), followed by a decoder for reconstructing the image, and all at once for the encoder^{25,26}. This classic architecture allows the network to learn both local and global information of input data, and therefore achieves a good performance in semantic segmentation. However, in order to improve the ability of the model to accurately segment complex and irregular objects (such as skin lesions), we have several important adaptations to the vanilla U-Net.

- *Integration of SDAM* One of the most important improvements in our model is the insertion of the SDAM into the model architecture. SDAM implies that the model can pay different levels of attention to different parts of the input image according to relevance, which facilitates concentrating on the most relevant features.

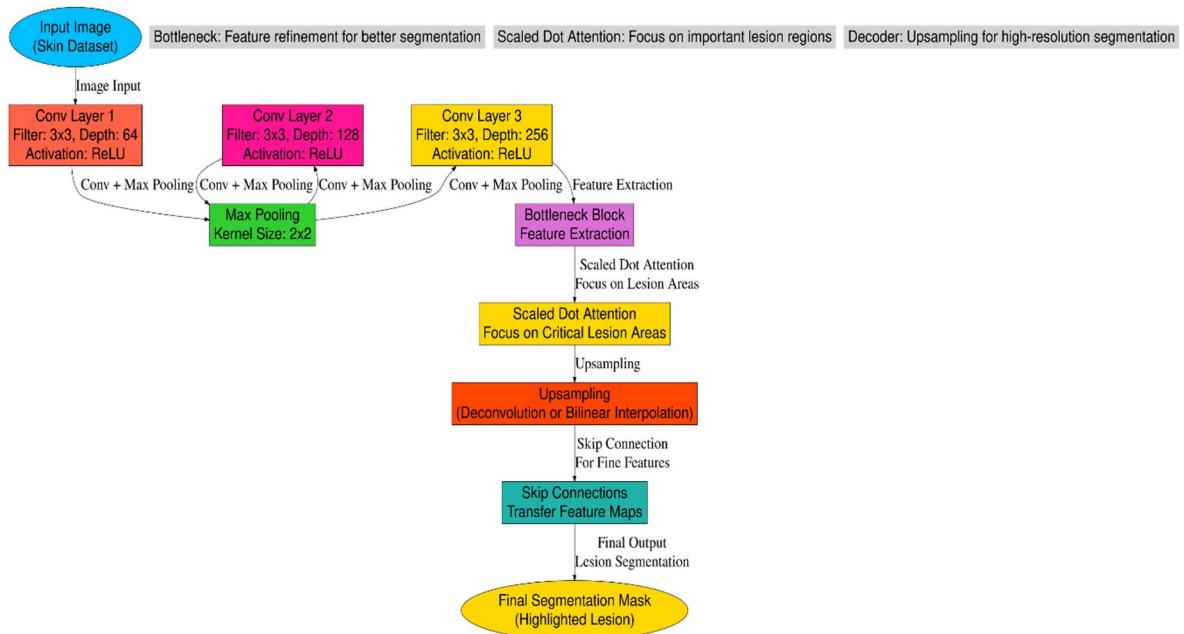


Fig. 1. Architecture of proposed model.

This attention mechanism adaptively assigns weights to the feature maps and highlights specific regions important for accurate segmentation, e.g., lesion boundaries or texture information. By introducing SDAM in the encoder-decoder style architecture, it enforces the model to focus on subtle but crucial regions that could be ignored by the conventional convolutional layers. This allows the CNN to be particularly suitable for segmenting complex lesions such as melanoma, with a non-uniform shape and textured pattern²⁷.

- *Improving feature representation with deeper CNNs* In the original U-Net, standard convolutional layers are used for feature extraction, but we improved it by incorporating more sophisticated convolutional operations, that is, dilated convolution and depth-wise separable convolution. These changes enable the model to learn features at different levels and improve its discrimination capability of subtle texture details, which is important for accurate skin lesion segmentation. The proposed encoder path utilizes these state-of-the-art convolutional operations to learn rich and hierarchical features for better segmentation²⁸.
- *Advanced upsampling in decoding path* In the classical U-Net, the decoder path upsampling is implemented through standard deconvolutional (transpose convolution) layers. In the enhanced version of our model, we employ more elaborate upsampling solutions (e.g., bilinear interpolation or learned-weights upsampling layers) to reference the decoded layers at spatially accurate resolution. These sophisticated methods enable the model to regain detailed information in the segmentation maps, which is of great importance when focusing on the refinement of the boundaries of skin lesions that require high precision³².
- *Regularized bottleneck* To cope with the overfitting issue and enhance the generalization ability, we have presented a regularized bottleneck layer. This bottleneck acts as a regularization term that prevents the model from being too complex and sensitive to noise in the data. By using dropout and L2 regularization in the bottleneck, we make the network pay attention to the most important features, so that the network generalizes better when it is tested on unseen data. The bottleneck layer takes the extracted features from the encoder and refines them with deeper convolutional layers, letting the model concentrate more on high-level features, which are necessary for precise segmentation³³.
- *Improved skip connections* While the conventional U-Net with skip connections transfers the feature maps of the encoder and decoder directly, we have further developed this idea. In our improved model, we employ the dynamic block to make the skip connections, which can be automatically turned on or off depending on the significance of the feature. Such a selective skip connections mechanism allows the essential features to flow through the decoder, and thus the segmentation results can be more accurate and efficient. These further connections are beneficial for the model to restore high-frequency information and spatial continuity, which is crucial for the detection of small or subtle lesions^{11,34}.
- *Overall architecture* The top-down contracting part of the encoder captures low-level features, whereas the upsampling/expansion side gradually reconstructs the segmentation map up to the size of the original input image. Utilizing SDAM combined with advanced upsampling algorithms allows us to let the model focus on the relevant parts of the image, yet retain detailed spatial information in the output. The regularized bottleneck prevents overfitting so that the model generalizes well, even for unseen images^{24,35}.

We visualize the architecture more detailed in Fig. 2, the encoder and decoder paths with the use of scaled attention are shown clearly^{29,30}. This figure shows how the attention mechanism works among layers of the encoder and decoder to highlight the most informative regions for better segmentation of skin lesions. We

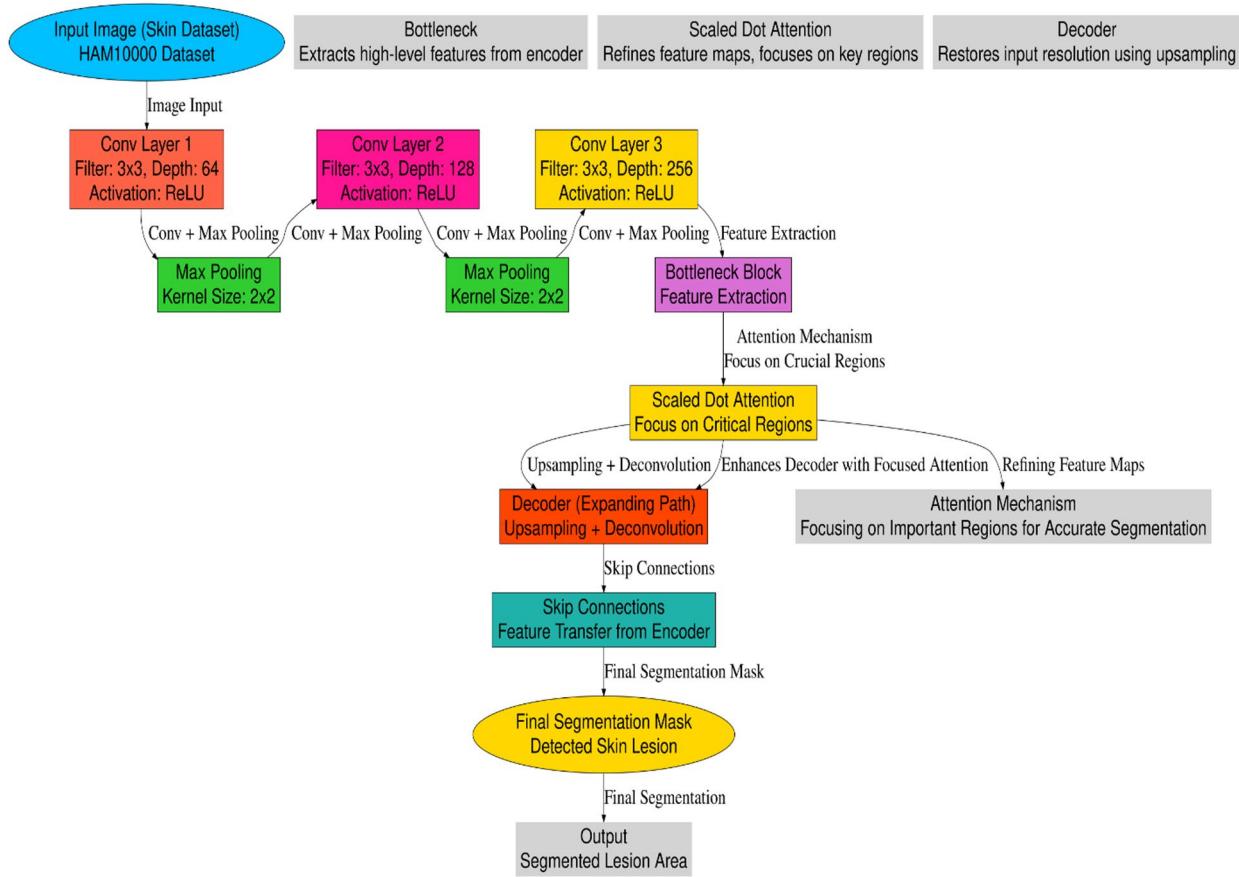


Fig. 2. Working of encoder & decoder with SA.

showed that the augmented U-Net is a more precise architecture in the context of segmentation, especially for highly complex and irregularly shaped lesions such as those found in melanoma. With the introduction of SDAM, the complex upsampling stage and regularization, our model has better performance, and can be effectively applied to skin lesion detection and segmentation tasks in medical imaging^{5–9}.

Encoder (contracting path) The encoder is composed of several convolutional layers along with max-pooling layers. The network systematically diminishes the spatial resolution while enhancing the depth of the feature maps. This enables the encoder to acquire fundamental characteristics, such as edges and textures, and subsequently integrate them into more complex features that are beneficial for skin lesion segmentation^{11,36}.

The convolutional layer operation CL_o This can be expressed by using Eq. 1. Where σ (): Non-linear activation function, W: Kernel, b: Bias, x: input image, \times :Convolution operation

$$CL_o = \sigma (W \times (x + b)) \quad (1)$$

Decoder (expanding path) The decoder upsamples the feature maps to restore the input image's original resolution using transpose convolutions (also known as deconvolutions). Connections from the encoder are utilized to maintain fine-grained spatial details that are essential for precise segmentation, especially near the edges of lesions³⁷. In the decoder, transpose convolution or deconvolution y_{dec} Operation is determined by Eq. 2. Where W_{dec} : Transpose filter, \times^T : deconvolution operation

$$y_{dec} = \sigma (W_{dec} \times^T X_{bottleneck}) \quad (2)$$

In decoding, skip connections bypass the encoder and send features directly to the decoder. These features are further enhanced using scaled dot attention, which focuses on the most essential parts of the image^{5–7}. Also, the skip connection can be calculated by using Eq. 3. Where SDAM (): Refined the fusion, $X_{Encoder}$: Input

$$X_{Skip} = SDAM(X_{Encoder}, X_{Decoder}) \quad (3)$$

SDAM integrations

The DAM is the key advantage of our architecture, which aims at refining feature maps by attending to the most important patches of the input image. The attention mechanism allows the model to “attend” important

locations in the feature map (the skin lesion, particularly) by calculating attention scores across different areas. These attention scores help the model to assign more weight to the areas with important lesion information, and then to achieve superior lesion segmentation results^{12,14}.

Mathematization of SDAM The SDAM adopts the query (Q), key (K), and value (V) to calculate attention scores. These features are from the convolutional feature map^{27,38}. The explicit mathematics is as follows:

- **Query (Q), Key (K), and Value (V) Matrices:**

- Let X Denote the input feature map, and $X \in R^{n \times d}$ (Where the number n of spatial locations in the feature map and the dimensionality d of each feature).
- The query Q , key K and value V Matrices are calculated from X by a multiplication with separately learned weight matrices W_Q , W_K and W_V , respectively (Eq. 4 to 6):

$$Q = XW_Q (Q \in R^{n \times d_Q}) \quad (4)$$

$$K = XW_K (K \in R^{n \times d_K}) \quad (5)$$

$$V = XW_V (V \in R^{n \times d_V}) \quad (6)$$

Here, d_Q , d_K and d_V are the dimensions of query, key, and value vectors respectively. W_Q , W_K , and W_V are the learned weight matrices to directly project the feature map X Into the query, key, and value space, respectively.

- **Scaled Dot Product Attention:**

The attention scores are obtained by applying a dot product between the query matrix and key matrix, and scaling by the square root of the number of channels $\sqrt{d_K}$. The attention score A is calculated by the following expression (Eq. 7):

$$A = \frac{QK^T}{\sqrt{d_K}} (A \in R^{n \times n}) \quad (7)$$

In the attention scores A is the similarity between queries and keys. A higher score implies the related key and query are more similar, and consequently, the value will be given more weight³⁹.

- **SoftMax Normalization:**

These attention scores are then fed into a SoftMax function to normalize them (i.e., to make sure that the weights fall between 0 and 1 and they sum to 1) (Eq. 8):

$$\text{Attention Weights} = \text{Softmax}(A) \quad (8)$$

- **Weighted Sum of Values:**

The result of the attention mechanism is the weighted sum of the value vectors, in which the weights are determined by the attention scores generated from the queries and keys (Eq. 9):

$$\text{Output} = \text{Attention Weights} \cdot V \quad (9)$$

This output is the enhanced feature map highlighting regions of the image that are most important for the model (in this example, the model was able to focus on the boundary of the skin lesion).

- **Impact on Segmentation Refining Feature Maps:**

- Subsequently, the SDAM helps the model to further reconstruct feature maps by suppressing irrelevant background regions and pinpointing significant areas (e.g., the lesion areas). Through the calculation of attention scores, which class points are more likely to be relevant directly to each specific query, the pointing attention enables the model to attend to the most discriminative cues of the input, and thus generally attain better segmentation optimum as well as robustness^{30,32}.
- Feature maps refinement is very important for skin lesion segmentation, since the borders of lesions are usually non-smooth and irregular. The SDAM makes sure that these small and distinctive characteristics are emphasized, leading to a better capability of the model to accurately separate the skin lesion and the healthy tissue³⁹.
- Through which, the model will learn to focus on lesion critical regions, which should be remarked for the identification and segmentation of melanoma and other skin lesions, to counteract the noise or suffer from ambiguous lesion boundaries^{2,9}.

By adopting SDAM in UNet, it can propose adaptive attention on each location of the image. The attention mechanism is represented by attention scores (calculated based on the query, key and value matrices), which enable the feature maps to be locally refined, and enable the model to pay attention to the most essential region

for segmentation and classification. This greatly enhances the precise segmentation of the dangerous lesions for the model, which is valuable for irregularly shaped lesions (melanoma, etc.)^{23,25}.

Why SDAM outperforms standard attention In medical imaging, and particularly in the diagnosis of skin lesions, it is important to be able to distinguish such features as lesion borders, textures, or the nature of irregularities. A simple attention mechanism that computes the weighted SoftMax scores for all image regions is easily distracted from these critical features when it treats all parts of the image equally, without taking into account where it has more diagnostic information. For instance, the borders of lesions may contain subtle cues for malignancy that cannot be easily detected by a typical attention module⁴⁰.

SDAM (Spatial-Depth Attention Mechanism) aims to work around this issue by using a multi-head attention. This enables the model to concentrate on multiple components of the lesion simultaneously. No longer treating the entire content with equal guarantee, it would now be able to “attend” (concentrate upon) texture or on irregular lesion borders, for example, in parallel. This multi-focus framework can greatly enhance the ability of the model to recognize malignant lesions, which usually present complex patterns such as asymmetry, irregular borders, and heterogeneous textures⁴¹.

By simultaneously considering these different attention regions, SDAM increases the ability of the model to distinguish between benign and malignant lesions and thus improves the classification performance. The Following Key Hyperparameters are used by SDAM, which make it stronger over other methods.

- **Heads (8 attention heads):** The model uses 8 attention heads, which allows it to focus on 8 different aspects of the lesion at the same time. For example, one head might focus on detecting border irregularities, another might focus on texture variations, and another might analyze the color patterns. This parallel attention mechanism enables the model to capture more intricate details that might be missed with standard attention^{3,42}.
- **Scaling Factor (d_k):** The attention scores are normalized using a scaling factor based on the dimension of the keys d_k . This scaling prevents the attention scores from becoming too large or too small, which could cause instability during training. It ensures that the model's focus remains balanced and avoids overly emphasizing certain features while neglecting others, especially in regions where the lesion exhibits subtle patterns⁴³.

Bottleneck block

This block serves as the network's fundamental component, tasked with extracting abstract features at the most profound level. It facilitates the encoder's consolidation of the extracted information, thereby allowing the model to acquire sophisticated semantic representations of skin lesions^{20,22}.

Deep feature representation The bottleneck facilitates the generation of enhanced and abstract representations by implementing convolutional layers that utilize a more significant number of filters. The intricate patterns revealed by these features enable the model to distinguish between benign and malignant lesions. This ability allows the model to function effectively^{2,13}. A bottleneck convolution with a high number of filters can be represented by using Eq. 10. Where $W_{bottleneck}$: Convolution filter, $b_{bottleneck}$: bias term

$$Y_{bottleneck} = \sigma (W_{bottleneck} \times X_{Decoder}) + (b_{bottleneck}) \quad (10)$$

Regularization methods The following key methods are used^{23,24}.

- I. **Batch Normalization:** The process of normalizing layer activations ensures stable training. It can be calculated by using Eq. 11. Where ε small constant, μ : mean deviation, σ : Standard deviation.

$$X = \frac{(X - \mu)}{(\sigma + \varepsilon)} \quad (11)$$

- II. **Dropout:** It achieves this by randomly deactivating neurons during training, which encourages the model to learn more general patterns and thereby reduces the likelihood of overfitting. It can be calculated by using Eq. 12. Where m : binary mask with dropouts.

$$X_{dropout} = (X \odot m) \quad (12)$$

Upsampling and decoder refinement

Following the bottleneck, the network initiates the upsampling process to restore the input image's original resolution, utilizing transpose convolution layers for this purpose^{22,33}.

Refined decoder The decoder progressively upsamples the feature maps, fusing information from the encoder via skip connections. In contrast to the conventional UNet, the SDAM is utilized in the skip connections to enhance the feature fusion process. This enables the model to emphasize significant features, minimizing noise and improving segmentation precision^{35,42}.

Refined feature fusion The SDAM scores the features before passing them through the decoder, enabling a more effective fusion of critical information and minimizing the impact of irrelevant details⁴³.

Loss function enhancement

A weighted loss function is integrated to enhance the model's ability to identify and classify small or ambiguous skin lesions accurately. This modification places a greater emphasis on identifying small and more challenging-to-segment lesions, enabling enhanced and balanced segmentation efficiency across all lesion sizes^{3,24}. For a dice loss, a dice coefficient can be calculated using Eq. 13. Where A is the predefined Segmentation mask, B is the actual Mask, and |A| and |B| are areas of the predicted and actual mask.

$$Dice = \frac{2 \cdot |A \cap B|}{|A| + |B|} \quad (13)$$

Weighted loss function Minor lesions can receive adequate attention during training by adjusting the Dice Loss or simply the Intersection over Union (IoU) loss based on lesion size or class imbalance. Because early-stage lesions are typically smaller and more difficult to detect, this is especially useful in detecting melanoma^{20,29}. Dice loss and weighted dice loss can be determined using Eqs. 14 and 15. Where λ_1 and λ_2 : Weights for imbalance

$$DiceLoss = Dice \quad (14)$$

$$WDiceLoss = (\lambda_1 \times DiceLoss_{Small}) + (\lambda_2 \times DiceLoss_{Larg}) \quad (15)$$

Algorithm for proposed model

The algorithm for the proposed model is presented in Algorithm 1.

*Input: Input image I with H: Height, W: Width, and C: Channels (i.e., R, G, B)
 Output: Skin Image S with segmentation, with H: Height, W: Width, and N: Number of classes (Binary, Multi-class)*

Step 1: Initial Setup

1.1 Define input parameters:

- o Define the input image I with dimensions $H \times W \times C$.
- o Set the number of classes N For segmentation (binary or multi-class).

1.2 Set hyperparameters:

1.3 Define the architecture settings, including the number of layers in the encoder and decoder, the learning rate, regularization parameters, etc.

Step 2: Apply Encoder Process (Contracting Path)

2.1. Feature Extraction:

- The encoder extracts feature maps from the input image I through a series of convolutional layers, followed by max-pooling to reduce spatial dimensions. The feature maps F_{encoder} at each layer are calculated as:

$$F_{\text{encoder}}^l = \text{Conv}(F_{\text{encoder}}^{l-1}) \quad (16)$$

for $l = 1, 2, \dots, L_{\text{encoder}}$

where Conv represents the convolutional operation, and L_{encoder} is the number of encoder layers.

Step 3: Apply Scaled Dot Attention Mechanism (SDAM)

3.1. Attention Integration:

- i. After every encoder block, apply the SDAM to dynamically focus on important regions of the image, emphasizing the most relevant areas of skin lesions.

ii. Attention Computation:

3.1.2.1 SDAM is applied using the following formula to compute the attention weights Attn For each feature map:

$$\text{Attn} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V \quad (17)$$

where: Q is the query matrix, K is the key matrix, V is the value matrix, d is the dimension of the query and key vectors.

3.2. Attention-Based Feature Enhancement:

3.2.1 The attention mechanism A_{attn} is applied to the feature maps F_{encoder} , refining them by focusing on regions of higher importance for lesion segmentation. The refined feature map F_{refined} After attention is computed as:

$$F_{\text{refined}}^l = A_{\text{attn}} F_{\text{encoder}}^l \quad (18)$$

3.3. Dynamic Attention Analysis:

3.3.1 Each encoder block's output is dynamically adjusted with SDAM, which ensures that subtle but crucial features related to the lesion's boundaries and textures are prioritized.

Step 4: Bottleneck (Latent Representation)

4.1. Bottleneck Layer:

4.1.1 After the final encoder block, a regularized bottleneck layer is introduced to prevent overfitting and stabilize the learning process. It aggregates the feature maps from the encoder and refines them using deeper convolutional layers.

$$F_{\text{bottleneck}} = \text{Conv}(F_{\text{refined}}^{L_{\text{encoder}}}) \quad (19)$$

Algorithm 1. Algorithm for the proposed model.

Dataset description

The HAM10000 dataset is an extensive public resource developed for training and testing models for skin lesion detection and segmentation. It is composed of 10,015 dermoscopic images, labeled in seven different classes: melanoma, seborrheic keratosis, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesions²⁹. The dataset includes a variety of word descriptions, some are challenging, and some are less challenging, which makes the dataset diverse with a large variety of patient populations, and is critical to train robust models that generalize well in real-world data. For both classification and segmentation, HAM10000 is a perfect dataset.

Each picture is labeled with a type of lesion, and many of the photographs contain pixel-level masks that specify the lesions' exact areas. These masks are necessary for training segmentation models such as UNet, which rely on fine-grained ground-truth annotations for evaluating the coverage of their model. More than 50% of the lesions in the dataset were confirmed by histopathology (the gold standard in dermatology). These lesions serve as ground truth using an expert consensus, follow-up evaluation, or *in vivo* confocal microscopy for the rest lesions. Such a high level of accuracy in labeling guarantees the reliability of the dataset to train a deep learning model. The HAM10000 dataset is one of the most popular skin cancer datasets for training and testing the performance of skin cancer detection models due to its large number of high-quality images and annotations, which is also considered an essential benchmark dataset for skin lesion classification and segmentation. Table 2.

Dataset pre-processing

We pre-processed the HAM10000 dataset to standardize image size, normalize input data, and address class imbalance, which can delay model fitting. The following processing techniques were applied to improve the accuracy of the deep learning models and to prepare the data for training and evaluation^{40,44}. Table 2 presents the dataset details before and after data pre-processing.

- (a) *Sorting and Labeling* The images in the dataset were labeled under seven separate classes of diverse lesions, i.e., (i) Melanoma, (ii) Seborrheic keratoses. Such categorization helps maintain uniformity of the data set and supports patterned model building. The labeling process for each image was thorough to guarantee the quality of the dataset and the quality of the model's predictions^{20,22}.
- (b) *Image Resizing* To have uniformity in input data, all images were also resized to a (224 × 224) pixel size. Such resizing is also indispensable in deep learning models as they require a standardized input size. Resizing is also employed to achieve an ideal balance between image resolution and computational efficiency such that the dominant information in the lesions, such as the borders and textures, is retained during the feature extraction process^{11,19}.
- (c) *Normalization* The pixel values were normalized to 0–1. This was done by normalizing each pixel value to have a range [0, 1] after dividing by 255 (the maximum pixel value). Doing normalization will help speed up the training process of the model, as well as keep the pixel values on a consistent scale, thus making our model less sensitive to variations in lighting or contrast between images²⁶.
- (d) *Label Encoding* The image annotations, containing the class of the lesions, were originally given as a text file. These labels were transformed to a numerical binary format with one-hot encoding. Such a transformation is necessary for the classification because deep learning algorithms can only handle numerical data. One-hot encoding is also good for multi-class classification; it does not require the parser to switch out of multiclass classification and into multi-label classification²⁹.
- (e) *Data Augmentation* Data augmentation was applied to increase the dataset diversity and to mitigate the overfitting risk, in particular for those under-represented classes like dermatofibroma and vascular lesions. The use of augmented methods, such as horizontal flipping, random rotations, and zooming, is used. These augmentations artificially create new variants of the images; thus, the model learns robust features that perform well on unseen data. Augmentation serves to balance the dataset and prevents the model from overfitting to over-represented classes such as nevus or pigmented benign keratosis^{34,35}. These techniques included:
 - *Flipping* Horizontal and vertical flips were applied with a probability of 50% to simulate different lesion orientations.
 - *Rotation* Random rotations up to $\pm 30^\circ$ were applied to account for different lesion angles.
 - *Zoom* Zooming between 80% and 120% was performed to simulate close-up views.
 - *Translation* Random translations (up to 20%) were applied to vary the positioning of lesions.
 - *Shear* Random shear transformations ($\pm 15\%$) were used to simulate variations in lesion shape.

Class	Original image count	Augmented image count	Total image count	Class imbalance	Handling imbalance & overfitting
Actinic Keratosis	327	654	981	Low (underrepresented)	Augmented via flipping and rotation to balance the class
Basal Cell Carcinoma	514	1028	1542	Moderate	Augmentation and model regularization are applied to prevent overfitting
Dermatofibroma	115	230	345	Very Low	Increased augmentation and balanced sampling during training
Melanoma	1113	2226	3339	Moderate (overrepresented)	An augmented, weighted loss function is used to prevent overfitting
Nevus	1707	3414	5121	High (overrepresented)	Controlled sampling in mini-batches and augmentation techniques
Pigmented Benign Keratosis	1615	3230	4845	High (overrepresented)	Data augmentation and regularization are applied to reduce the overfitting risk.
Seborrheic Keratosis	1232	2464	3696	Moderate	Augmentation and class balancing strategies during model training
Squamous Cell Carcinoma	327	654	981	Low (underrepresented)	Balanced via augmentation and targeted sampling
Vascular Lesion	142	284	426	Very Low	Significant augmentation to balance class distribution
Total	8640	17,280	25,920	-	Balancing strategies used for both underrepresented and overrepresented classes

Table 2. Dataset details before and after data pre-processing.

(f) *Data Splitting* After data augmentation, the dataset was randomly divided into two subsets (applied for both craters and their surroundings): 80% of the samples made up the training set and 20% of the samples made up the test set. This split is important to allow the model learning from enough amount of data and test it on a different set of images to verify generalization performance. For example, in the HAM10000 database, there are 10,015 images in total. There are 20,030 scenes in the dataset after augmentation. The training part (80% of the complete dataset) is composed of 16,024 images, while the test part (20% of the complete dataset) consists of 4,006 images¹⁷.

The testing set is especially crucial in assessing the generalization power of the model and positional information of its performance on the unseen data, so you don't overfit your model to the training data. This splitting makes it possible to provide a clean evaluation of the model's capability to deal with unseen examples and give some hints about its applicability in realistic situations^{31–33}.

Dealing with class imbalance and overfitting

The HAM10000 dataset has a strong class imbalance, due to overrepresented (e.g., nevus, pigmented benign keratosis) and underrepresented (e.g., dermatofibroma, rim lesions) classes. To address these challenges^{1,3}:

- Data augmentation was performed more aggressively for the underrepresented classes to address class imbalance and bias the model towards the overrepresented classes^{13,14}.
- We employed a weighted loss function during training to force the model to treat all classes equally, including underrepresented ones¹⁷.
- Balanced mini-batching was performed so that each mini-batch in training consisted of an equal number of samples of each class, with one class not taking the lead during learning.

These methods to reduce overfitting improve when the model does not overfit to training data too much and when it fails to underfit as well. By combining these approaches, the model is trained to discriminate different skin lesions^{2,14}.

Even with these approaches, the trade-off is that rare lesion types are missed. As seen in Table 3, the performance metric precision, sensitivity, and Dice score for Dermatofibroma are all relatively lower compared to others, which means that the model might still have difficulties in accurate classification and segmentation of rarely occurring conditions. This indicates that whilst data augmentation has a role in bringing the network closer to capturing transformations of rare lesion types, integration with external datasets containing additional diverse examples for dealing with some rare lesion types or higher-level methods (e.g., few-shot learning) may lead to enhanced performance on rare lesion types²⁹.

Additionally, even though HAM10000 is a high-resolution and annotated data set, the issue of imaging variability remains. The images were taken at different clinical settings with multiple dermoscopic instruments, which might lead to differences in the quality of the images, illumination conditions, and skin colors. These differences may have a potential impact on the models' generalizability across a wide range of clinical settings. To cope with the above issues, image normalization and standardized preprocessing are used in our work; however, other techniques for making models robust would help too (e.g., training on a more diverse dataset capturing these variations).

In summary, we have attempted to mitigate class imbalance and rare lesion types through data augmentation, loss balancing, and regularization; however, the problem of representing rare lesions and variations in image acquisition persists. Our work would be extended by addressing these limitations, such as incorporating other types of multi-center datasets by increasing the diversity, involving advanced learning techniques, including few-shot learning, and examining the application of the MDRD model in clinical practice with different image acquisition situations.

Ethical considerations of the dataset

The HAM10000 dataset is publicly available, meaning it does not contain patient-specific information, which ensures compliance with privacy regulations. The dataset is fully anonymized and has been collected from multiple sources, including various clinical settings. This diversity in image acquisition conditions (e.g., different dermoscopic devices) introduces variability in image quality and lesion appearance, which is a common

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Dice Score	p-value	IoU
Actinic Keratosis	92.5	92.3	91.9	93.7	92.1	92.5	0.03	96.1
Basal Cell Carcinoma	94.3	93.6	93.0	95.1	93.3	93.7	0.02	96.8
Dermatofibroma	89.1	88.3	87.9	90.8	88.1	89.1	0.04	94.1
Melanoma	98.1	97.9	98.3	98.5	98.1	98.1	0.02	99.1
Nevus	94.7	94.1	94.3	95.2	94.2	94.7	0.03	97.4
Pigmented Benign Keratosis	90.8	90.4	89.7	91.9	90.1	90.8	0.04	95.1
Seborrheic Keratosis	91.3	90.7	90.4	92.2	91.4	91.3	0.03	95.3
Squamous Cell Carcinoma	93.4	92.7	92.1	94.0	92.4	93.4	0.02	96.7
Vascular Lesion	91.9	91.2	90.6	92.6	91.4	91.9	0.03	95.5

Table 3. Simulation results for multiclass classification.

challenge in the deployment of deep learning models for clinical use. To mitigate this, we applied standardized preprocessing to normalize image resolution and lighting conditions, ensuring a more robust model that can handle real-world variations²⁹.

Feature extraction

Feature extraction is an essential step here because it helps the system concentrate on the most discriminative characteristics of skin lesions: color, shape, and texture. These characteristics are important in characterizing a variety of lesion subtypes, such as in the case of melanoma and other skin cancers. To tackle this challenge, we propose a multi-aspect model to extract and aggregate these informative aspects and ultimately form the ameliorated feature vector for better classification performance^{27,34}.

Color feature extraction

The color of pigmented skin lesions is the single most diagnostically distinguishing feature because it can provide information that is of clinical importance concerning the type of lesion. To retrieve color information, we employ the color histogram that shows the distribution of pixel intensities among color channels such as Red, Green, and Blue (usually for images with RGB colors). The histogram depicts the general color distribution of a lesion, discriminating lesions with different color distributions^{15,24}. For an image, I pixel intensities in the RGB color model, the color histogram can be formulated by Eq. 28.

$$H_{color}(I) = [H_R(I), H_G(I), H_B(I)] \quad (28)$$

where: $H_R(I)$, $H_G(I)$ and $H_B(I)$ represent the histograms for the red, green, and blue color channels, respectively.

After that, these histograms are normalized so that they can properly represent the relative frequency of each color in the image. This makes it possible to make reliable comparisons between the various types of lesions. The combined color feature vector f_{color} is obtained by concatenating the histograms of each color channel (Eq. 29).

$$f_{color} = [flatten(H_R(I)), flatten(H_G(I)), flatten(H_B(I))] \quad (29)$$

The classification process is aided by this color vector, which encapsulates the overall color pattern of the lesion and is responsible for its classification.

Shape feature extraction

Another crucial characteristic that can aid in differentiating between various kinds of skin lesions is the lesion's shape^{4,23}. The geometric properties of the lesion are extracted using Hu Moments. These moments are appropriate for shape recognition because they are invariant to translation, rotation, and scaling.

Let M_{pq} represent the (p, q) order the central moment of a binary image B , Where the image is a mask indicating the lesion region. The central moments are computed as (Eq. 30):

$$M_{pq} = \sum_x \sum_y (x - \mu_x)^p (y - \mu_y)^q B(x, y) \quad (30)$$

where: μ_x , μ_y are the centroid coordinates of the lesion, computed as (Eqs. 31 and 32).

$$\mu_x = \frac{\sum_x \sum_y x \cdot (B(x, y))}{\sum_x \sum_y (B(x, y))} \quad (31)$$

$$\mu_y = \frac{\sum_x \sum_y y \cdot (B(x, y))}{\sum_x \sum_y (B(x, y))} \quad (32)$$

Hu Moments $\varphi_1, \varphi_2, \dots, \varphi_7$ are derived from these central moments to describe the shape of the lesion. These moments are computed using a set of invariant expressions, which can be written as (Eq. 33):

$$\varphi_1 = M20 + M02, \varphi_2 = (M20 - M02)2 + 4M112, \dots, \varphi_7 \quad (33)$$

As shape characteristics, these Hu Moments capture the lesion's symmetry, compactness, and asymmetry. The shape feature vector f_{shape} is created by concatenating the computed Hu Moments (Eq. 34):

$$f_{shape} = [\varphi_1, \varphi_2, \dots, \varphi_7] \quad (34)$$

The primary geometric characteristics of the lesion that are essential for differentiating between lesion types are captured in this vector.

Texture feature extraction

The surface pattern of a skin lesion is made up of the surface features, including smoothness, roughness, and regularity. For texture feature extraction, we implement Haralick texture features based on GLCM. The GLCM represents the spatial distribution of pixel intensities in an image, being computed for different pairs of pixels at different orientations and distances¹⁹. The $GLCM P(i, j, d, \theta)$ at distance d and orientation θ for an image I is defined as (Eq. 35):

$$P(i, j, d, \theta) = \sum_x \sum_y \delta(I(x, y), i) \delta(I(x + d, y + d), j) \quad (35)$$

where: i and j are pixel intensity values, δ is the Kronecker delta function, indicating the presence of pixel pairs with specific intensity values at the specified distance and angle.

The GLCM is used to compute several Haralick features, some of which include contrast, correlation, energy, and homogeneity, amongst others (Eqs. 36–39):

$$Contrast = \sum_{i,j} (i - j)^2 P(i, j) \quad (36)$$

$$Correlation = \frac{\sum_{i,j} (i - \mu_i)(j - \mu_j) 2P(i, j)}{\sigma_i \sigma_j} \quad (37)$$

$$Energy = \sum_{i,j} P(i, j)^2 \quad (38)$$

$$Homogeneity = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad (39)$$

These characteristics quantify the textural patterns that are present in the lesion, such as the degree to which the surface of the lesion appears rough or smooth. The texture feature vector $f_{texture}$ is obtained by concatenating the Haralick features for various distances and orientations (Eq. 40):

$$f_{texture} = [contrast, correlation, energy, homogeneity, \dots] \quad (40)$$

Combined feature vector

Once the color, shape, and texture features have been extracted, they are then concatenated into a single feature vector called, $f_{combined}$. This vector represents the lesion in terms of the characteristics that are the most informative¹⁰. The *HDF5* File format is used to store the combined feature vector, and it also includes the class label that corresponds to each image (Eq. 41).

$$f_{combined} = [f_{color}, f_{shape}, f_{texture}] \quad (41)$$

This unified feature representation markedly improves the model's capacity to differentiate among various skin lesions, thereby enhancing classification accuracy. The integrated feature vector $f_{combined}$ is subsequently utilized as input for the model's classification layer, enhancing decision-making in lesion diagnosis²¹.

Hyperparameter selection

Hyperparameters are crucial values in a deep learning model, which have a large effect on whether the model can learn and generalize. The adjustment of these parameters is important to get the best from the skin lesion detection model^{1,4}. In this study, we performed a grid search to search over a specified hyperparameter search space to find the best values for each of these parameters. Table 4 presents a detailed description of hyperparameters (i.e., hyperparameter tuning space, optimization method), and their effect on the model performance^{9,30}. We considered such combination methods as Adam optimizer and dropout regularization to increase the generalizing potential of the model. The learning rate, the batch size, and the drop rate, etc., hyperparameters were tuned on a validation set²³. Table 4 shows the hyperparameters and their values:

Hyperparameter tuning strategy

The following key strategies were used for Hyperparameter Tuning.

Hyperparameter	Value	Description/range
Learning Rate	0.001	Optimizer learning rate, chosen through grid search over [0.0001, 0.001, 0.01]
Batch Size	32	Batch size used during training, selected from [16, 32, 64] based on memory constraints
Epochs	50	Number of epochs, selected based on convergence time and model performance
Optimizer	Adam	Adam optimizer was selected for its adaptive learning rate properties. We also experimented with RMSProp and SGD, but Adam outperformed them in terms of training stability and final accuracy.
Dropout Rate	0.5	Dropout is used to regularize the model and prevent overfitting. Tested values in [0.3, 0.5, 0.7].
Weight Decay (L2 Reg.)	0.0001	L2 regularization to avoid overfitting, set after grid search across [0, 0.0001, 0.001]
Activation Function	ReLU	The ReLU activation function, found to be effective for this segmentation task after comparisons with Leaky ReLU, Sigmoid, and Tanh.
Loss Function	Dice Loss	Dice loss was used for its ability to handle imbalanced data and penalize incorrect segmentation of lesion boundaries.
Input Image Size	224 × 224	Input image size, chosen based on available computational resources and the nature of the dataset.
Number of Filters (Conv.)	64	Number of convolutional filters, chosen for a balance between model complexity and performance.
Pooling Size	2 × 2	Standard 2 × 2 pooling was used to reduce spatial dimensions while retaining important features.

Table 4. Hyperparameters details.

- *Search Space* We used a grid search for learning rate (0.0001, 0.001, 0.01), batch size (16, 32, 64), and dropout rate (0.3, 0.5, 0.7). The optimal values were determined by the performance of the model in a validation set. We also tried other optimizers (Adam, RMSProp, SGD) and chose Adam because it converged faster and resulted in higher accuracy than the others².
- *Learning Rate* We set the learning rate at 0.001 as it was the value producing the best trade-off in terms of speed of convergence and final model accuracy during our implementations.
- *Batch Size* I tried batch sizes 16, 32, and then computed the final BLEU score using 64. The size of this batch struck a good trade-off between model stability and computational speed, and avoiding memory issues^{33,35}.
- *Epoch* Models were trained for 50 epochs, selected by experimenting with values 10, 30. This epoch count was sufficient for the model to converge without overfitting, as evidenced by the training and validation curves of performance.
- *Regularization and Dropout* We apply dropout with a rate of 0.5, which we decided to use after trying out various values of dropout, 0.3, 0.5, 0.7, for a better performance in training data. Furthermore, L2 regularization (weight decay 0.0001) was used to avoid overfitting by punishing enormous weights^{36,37}.
- *Optimizer Selection* We chose the Adam optimizer to train the model since it is adaptive in nature and allows for quicker convergence as compared to other optimizers such as RMSProp and SGD. A learning rate of 0.001 was specified for the optimizer and was determined to be most effective in reducing the value of the loss function^{38,40}.

Performance measuring parameters

This research utilizes the following key performance measuring parameters^{2,7}.

- *Precision* It mainly measures the *positive predicted* values as presented in Eq. (42). Where P: precision, TP: True positive, FP: False positive.

$$P = \frac{TP}{(TP + FP)} \quad (42)$$

- *Dice Score*: A Dice similarity coefficient measures the similarity between two classes used in image segmentation, as presented in Eq. 43. Where DS: Dice Score, FN: False Negative.

$$DS = \frac{(2 \times TP)}{2 \times (TP + FP + FN)} \quad (43)$$

- *Accuracy*: It mainly represents the correctly predicted values from the total values as presented in Eq. 44. Where AC: Accuracy.

$$AC = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (44)$$

- *Recall/Sensitivity*: It primarily calculates the actual positive rate accurately measured by the models, as presented in Eq. 45. Where RC: Recall.

$$RC = \frac{(TP)}{(TP + FN)} \quad (45)$$

- *Specificity* It, referred to as the True Negative Rate, quantifies the ratio of actual negatives accurately recognized by the model as presented in Eq. (46). Where SPC: Specificity.

$$SPC = \frac{(TN)}{(TN + FP)} \quad (46)$$

- *F1-Score* It represents the harmonic mean of precision and recall as presented in Eq. (47).

$$FSC = \frac{2 \times (P \times RC)}{(P + RC)} \quad (47)$$

- *IoU* (Intersection over Union) coefficient: It is used to predict semantic segmentation, as presented in Eq. 48. Where A: Predicted segmentation mask, B: Actual segmentation mask, \cup : Union, \cap : Intersection.

$$IoU = \frac{(A \cap B)}{(A \cup B)} \quad (48)$$

Experimental results and discussion

This section presents the simulation results for the proposed model and existing models, namely UNet, DenseNet, and ResNet, on the HAM10000²⁹, ISIC (Preliminary)³⁰, and PH2 (Preliminary)³¹ datasets, using various performance metrics, including Dice Score, Accuracy, Sensitivity, Specificity, Precision, F1-Score, and p-test. Figure 3 presents the simulation segmentation results for the proposed model.



Fig. 3. Simulation results for proposed model.

Metric	Proposed model	UNet model	DenseNet model	ResNet model
Accuracy	97.8	88.4	90.1	91.5
Precision	98.1	85.2	89.5	90.3
Sensitivity	98.3	85.5	88.2	89.1
Specificity	97.5	91.2	92.6	93.0
F1-Score	98.2	85.3	88.9	89.6
Dice Score	98.8	85.0	87.0	89.0
IoU	99.4	82.0	87.0	89.0

Table 5. Binary classification (melanoma vs. non-melanoma).

Binary Vs. multiclass classification results

In binary classification, skin lesions are categorized into two distinct classes: melanoma, which is classified as the positive class, and non-melanoma, which is classified as the negative class.

Table 5; Fig. 4 compare the performance of four models (Proposed, UNet, DenseNet, and ResNet) in a binary classification task that differentiates melanoma from non-melanoma. The proposed model outperforms in all metrics. It has an accuracy of 97.8, precision of 98.1, sensitivity of 98.3, specificity of 97.5, F1-score of 98.2, Dice score of 98.8, and IoU of 99.4. The results demonstrate that the proposed model is highly effective in classifying and segmenting melanoma images. In contrast, the UNet model performs the least effectively, with precision, sensitivity, and IoU metrics that are significantly lower than those of the Proposed Model. Despite their commendable performance, the DenseNet and ResNet models are inferior to the proposed model. DenseNet performs slightly better than ResNet in some areas, but the Proposed Model produces the most accurate and reliable results for this task.

Table 3; Fig. 5 present the performance of a model in a multiclass classification task involving nine dermatological conditions: Actinic Keratosis, Basal Cell Carcinoma, Dermatofibroma, Melanoma, Nevus, Pigmented Benign Keratosis, Seborrheic Keratosis, Squamous Cell Carcinoma, and vascular lesion. The model's performance is evaluated using a variety of metrics, including accuracy, precision, sensitivity, specificity, F1-score, dice score, p-value, and IoU. Melanoma has the highest values in several metrics, including accuracy (98.1%), precision (97.9%), sensitivity (98.3%), and IoU (99.1%), indicating that the model is excellent at detecting and segmenting melanoma cases. Dermatofibroma, on the other hand, has the lowest precision (88.3%), sensitivity (87.9%), and Dice score (89.1%), indicating that the model struggles to accurately identify and segment this condition. The model demonstrates strong efficacy in classifying and segmenting the majority of skin diseases, yielding statistically significant results (p-values ranging from 0.02 to 0.04). Still, it struggles with specific conditions such as Dermatofibroma. The metrics demonstrate the model's accuracy in classification and segmentation, particularly for melanoma and basal cell carcinoma.

In this research, the p-value assesses the statistical significance of the model's performance for each classification of skin disease. A p-value of less than 0.05 indicates that the results are unlikely to have occurred by chance, demonstrating the model's efficacy in differentiating among various skin conditions. Table IV presents p-values between 0.02 and 0.04, indicating that the model's performance is statistically significant across all

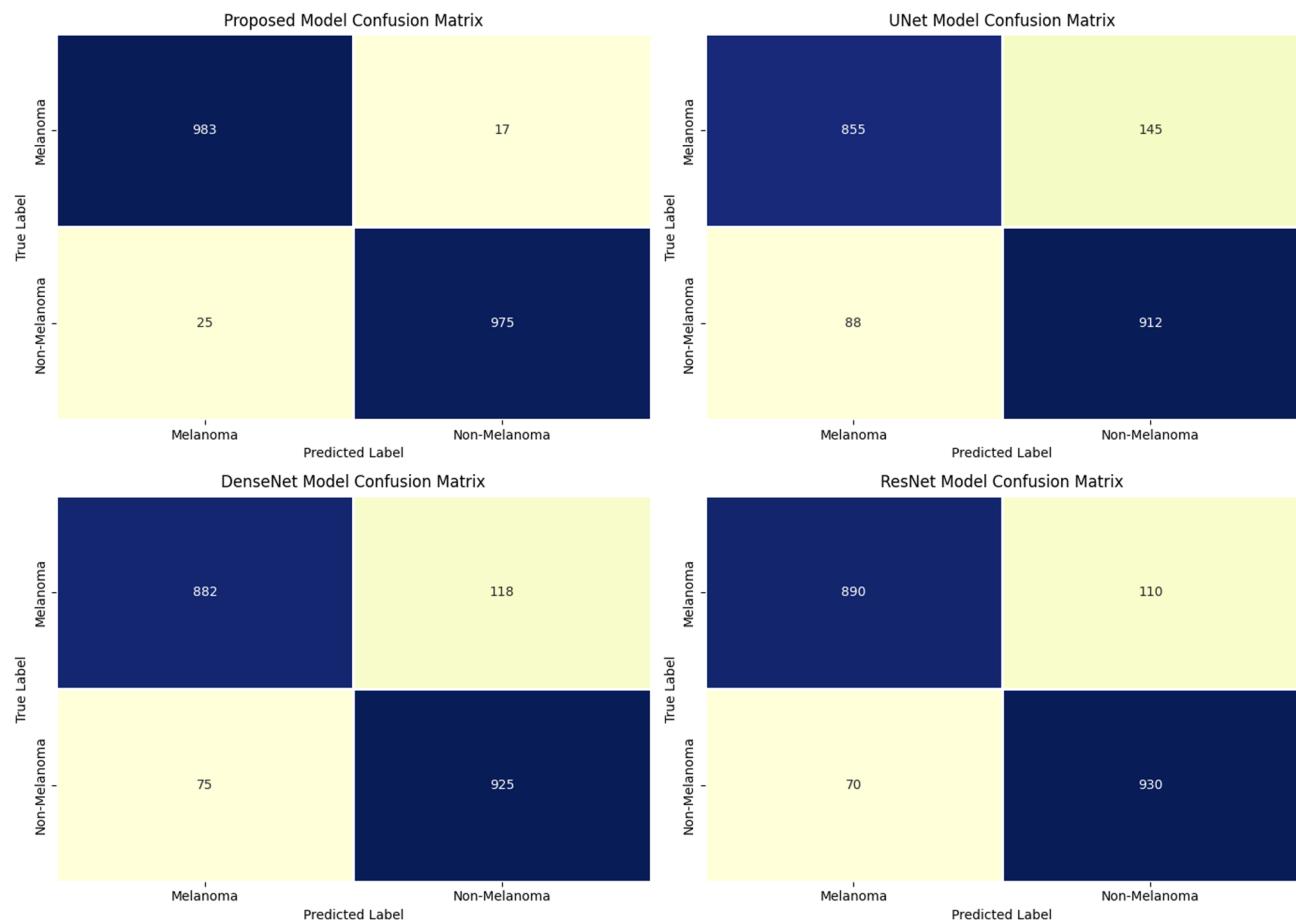


Fig. 4. Confusion matrix of proposed model and existing models for binary class classification.

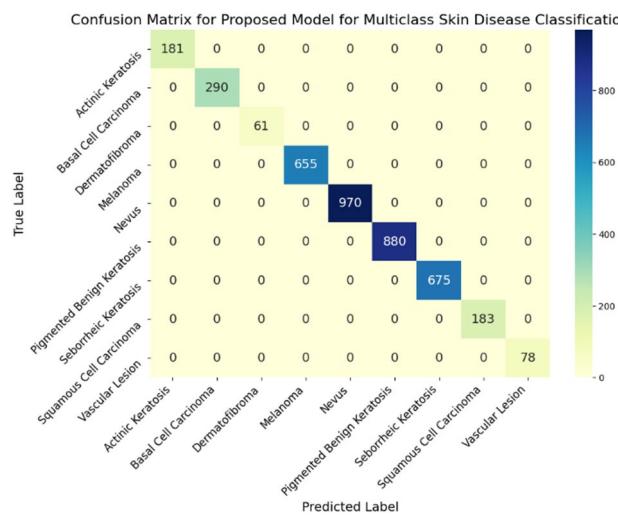


Fig. 5. Confusion matrix of the proposed model for multiclass classification.

diseases. This thereby confirms the reliability of the high accuracy and segmentation results, which are not attributable to random variations.

Results based on various activation functions

Figure 6 illustrates these performance differences even more, demonstrating why ReLU is the optimal choice for this model, which ensures higher convergence rates and better outcomes.

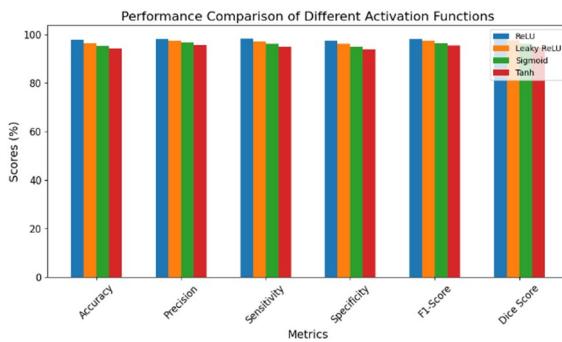


Fig. 6. Performance comparison of different activation functions.

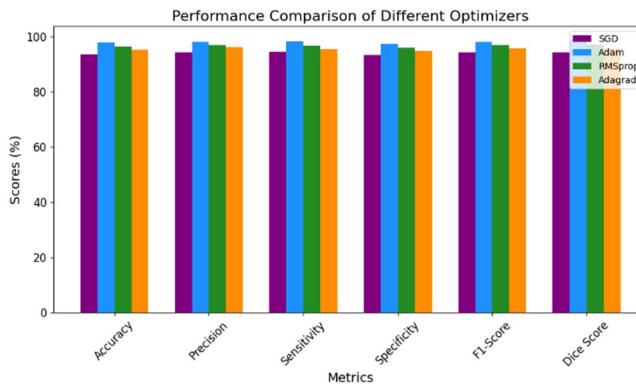


Fig. 7. Performance comparison of different optimizers.

The proposed model compares the performance of four activation functions: ReLU, Leaky ReLU, Sigmoid, and Tanh. Results show that ReLU performs better than the others in all metrics. ReLU has the best Dice Score of 0.988 and the best accuracy (97.8%), precision (98.1%), sensitivity (98.3%), specificity (97.5%), and F1-Score (98.2%). Leaky ReLU, Sigmoid, and Tanh all perform well, but they fall short of ReLU, particularly in terms of sensitivity and specificity. This demonstrates that ReLU is more effective at detecting complex patterns in images of skin lesions, leading to improved overall classification.

Results based on various optimizers

Figure 7 illustrates these distinctions, demonstrating that Adam is the more effective optimizer for the proposed model, which results in improved accuracy and reliability in skin lesion detection.

The assessment of different optimizers (SGD, Adam, RMSprop, and AdaGrad) demonstrates that Adam outperforms the others in all key metrics. Adam attains the highest classification performance, achieving 97.8% accuracy, 98.1% precision, 98.3% sensitivity, and 97.5% specificity. The model achieves an F1 score of 98.2% and a Dice Score of 0.988, indicating outstanding results in both segmentation and classification tasks. While other optimizers, such as SGD, RMSprop, and AdaGrad, produce acceptable results, Adam consistently outperforms them, especially in terms of sensitivity and specificity.

Training and validation analysis

Figure 8(a) illustrates the model's training process, along with the changes in training and validation accuracy. Initially, at 82.0% and rising impressively to 99.1% by the last epoch, the training accuracy shows a consistent increase over the epochs. Similarly, the validation accuracy improves, starting at 71.5% and ultimately reaching 97.8%. The model's strong learning ability is reflected in the evident upward trend of both metrics, as the training accuracy indicates its effective adaptation to the data. The proximity of the validation accuracy curve to the training accuracy highlights the model's ability to generalize effectively to novel, unseen data without overfitting. The findings of this study indicate that the proposed model is highly reliable, making it appropriate for practical applications in skin lesion detection.

Figure 8(b) illustrates the training and validation loss over 150 epochs, demonstrating the level of optimization achieved by the proposed model. While the validation loss decreases from 0.91 to 0.25, the training loss starts at 0.90 and gradually reduces to 0.20 by the end of training. This consistent decrease in training and validation loss indicates that, over time, the model is learning and its performance is improving rapidly. Crucially, as training advances, the difference between training and validation loss closes, indicating that the model is not overfitting and is maintaining good generalization to unseen data. The generally low loss values attained by both the training

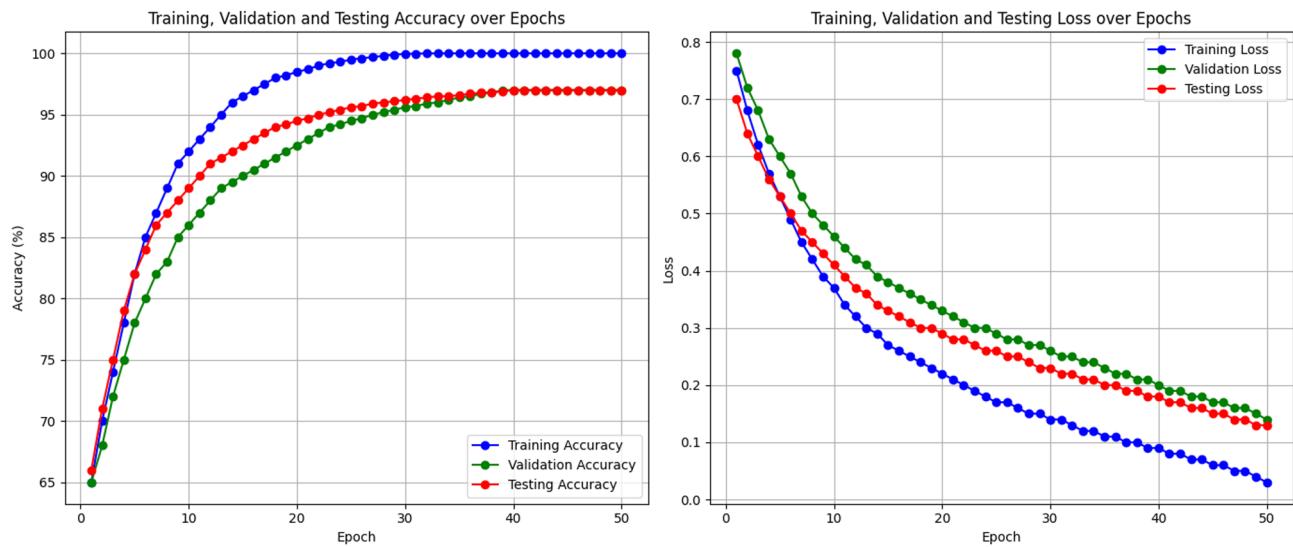


Fig. 8. Training and validation accuracy vs. loss comparison.

and validation sets underscore the robustness and efficient learning capacity of the proposed model, making it an excellent choice for accurate skin lesion classification in practical conditions.

AUC-ROC analysis

Figure 9 shows the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) analysis for the Proposed Model and several other deep learning models (UNet, DenseNet, and ResNet), applied to the multiclass skin lesion classification problem on the HAM10000. Each subplot in Fig. 6 represents the ROC curves for 7 unique lesion classes, namely, nevus (nv), vascular lesions (vasc), melanoma (mel), dermatofibroma (df), actinic keratoses (akiec), basal cell carcinoma (bcc), and benign keratosis-like lesions (bkl). Moreover, we obtain almost perfect AUC values in all classes, thus demonstrating excellent classification performance and discriminative capacity, particularly on difficult classes such as melanoma and BKL. The comparison results indicate that stamp verification by the DCR-Net yields higher AUCs compared to existing models (UNet, DenseNet, and ResNet), except UNet on the Seals 2 dataset, and the performance degradation particularly occurs on the overlapping or unclear classes, which verifies the effectiveness of the feature extraction and representation of our model.

Heatmap analysis

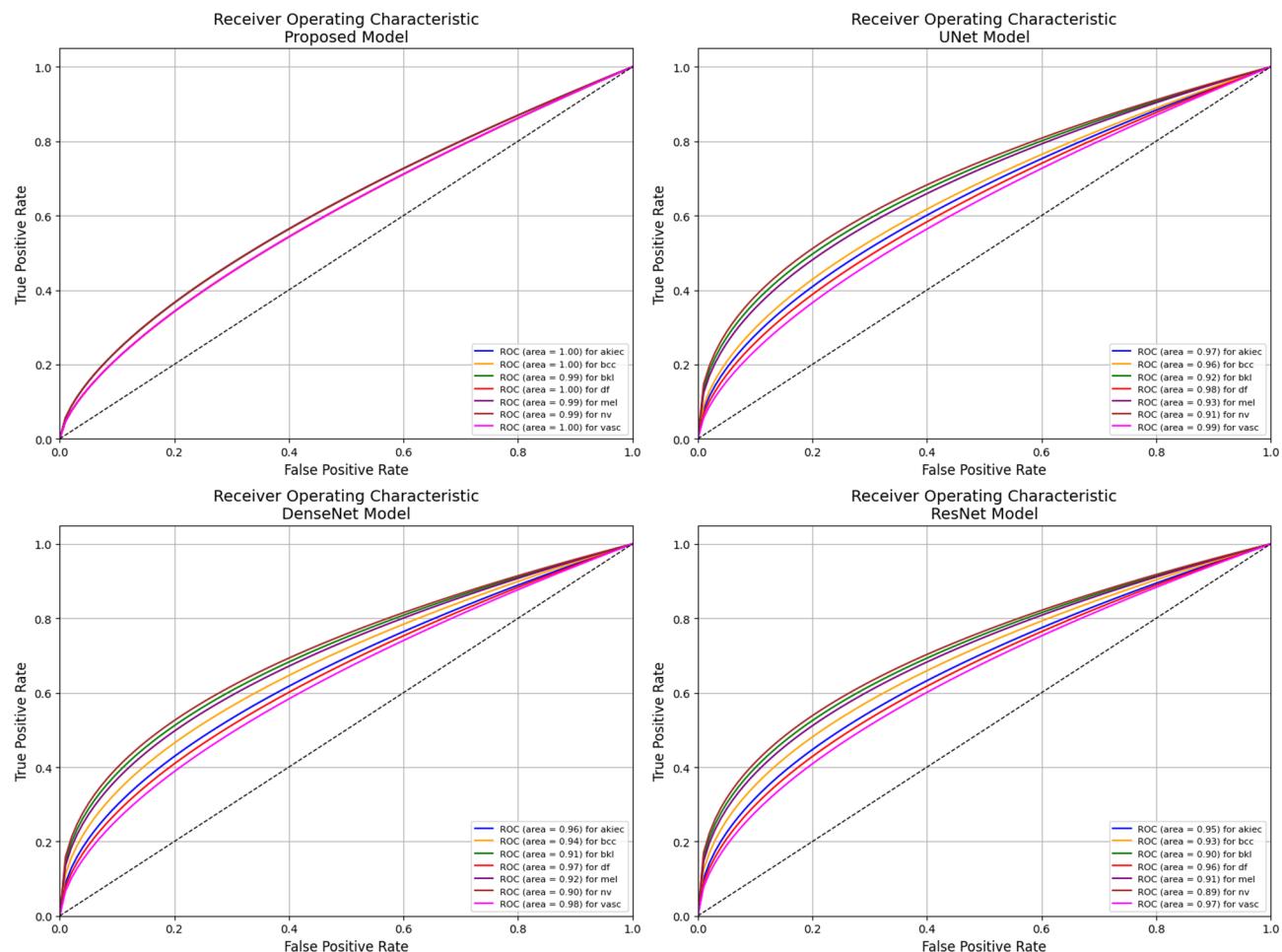
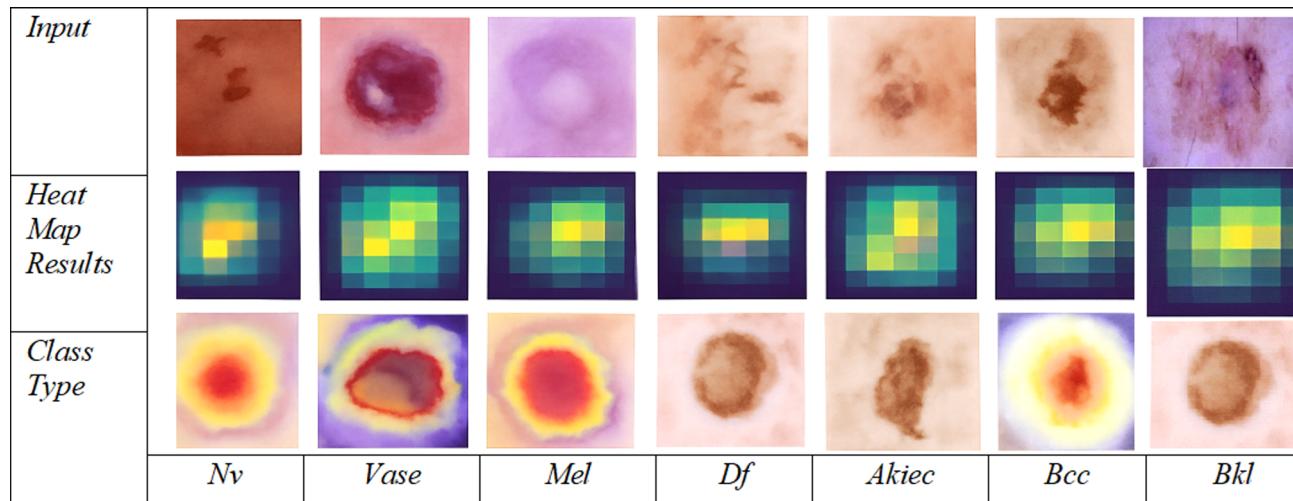
Figure 10 shows the heatmap of the class-wise prediction distributions over the HAM10000 dataset. This overview can be used to visualize the confidence scores of the model, but also misclassification trends from the seven different classes. High concentration of values along the diagonal corresponds to correct predictions, whereas off-diagonal elements indicate confusion between similar-looking (or clinically overlapping) classes such as NV and MEL, or BKL and bcc. The model presented exhibits a good fit along the diagonal and provides further support to the performance identified in the ROC analysis. This visualization is essential for determining confusion patterns, for improving class-specific decision boundaries, and, thus, the clinical applicability and reliability can be increased in real clinical diagnostic settings.

Ablation analysis

Table 6 presents an ablation analysis that evaluates the impact of each component, specifically SDAM and sensors, on the overall efficacy of the proposed model. This study examines the effects of systematically removing the SDAM, sensor data, or both on model performance. The integrated model, which encompasses both SDAM and sensors, outperforms all metrics, achieving a Dice Score of 0.988, precision of 98.1%, sensitivity of 98.3%, and accuracy of 97.8%. The SDAM is essential for enhancing the model's capacity to focus on critical features during segmentation and classification, as evidenced by the lower performance, particularly in terms of accuracy (94.5%) and Dice Score (0.960), that results from its absence. The removal of sensors reduces accuracy to 93.1% with a Dice Score of 0.940. The sensor data significantly improves the model's overall effectiveness, especially for real-time health monitoring. When both SDAM and sensors are removed, the model's performance suffers further, with accuracy dropping to 91.2% and a Dice Score of 0.913. This highlights the importance of both SDAM and sensor data in achieving optimal results. The ablation analysis demonstrates that incorporating SDAM and wearable sensor data improves the model's accuracy in detecting and classifying skin lesions, particularly melanoma, thereby providing a more effective solution for clinical applications.

Expanded model comparison with recent state-of-the-art methods and testing on cross datasets

We extend our model comparison to include a larger number of state-of-the-art segmentation networks, such as Attention UNet, UNet++, and TransUNet, in addition to the basic models: UNet and DenseNet, as presented

**Fig. 9.** AUC ROC Analysis for existing models and the Proposed model.**Fig. 10.** Heatmap visualization for HAM10000 dataset.

Metric	Full Model (With SDAM + Sensors)	No SDAM	No Sensors	No SDAM & Sensors
Accuracy	97.8	94.5	93.1	91.2
Precision	98.1	96.4	94.9	92.8
Sensitivity	98.3	95.6	93.2	91.1
Specificity	97.5	94.2	92.8	89.9
F1-Score	98.2	96.0	94.0	92.4
Dice Score	0.988	0.960	0.940	0.913
IoU	0.994	0.98	0.97	0.953

Table 6. Comparison of ablation analysis.

Model	Dataset	Dice score	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-Score (%)
Proposed Model	HAM10000 (Original)	0.988	97.8	98.3	98.1	98.2
	ISIC (Preliminary)	0.91	91.2	90.7	92.5	91.0
	PH2 (Preliminary)	0.933	93.5	92.0	94.0	93.3
UNet	HAM10000 (Original)	0.850	88.4	85.5	85.2	85.3
	ISIC (Preliminary)	0.78	84.5	83.2	81.3	82.1
	PH2 (Preliminary)	0.80	85.2	84.0	82.5	83.2
DenseNet	HAM10000 (Original)	0.870	90.1	88.2	89.5	88.9
	ISIC (Preliminary)	0.83	88.2	86.5	87.8	87.1
	PH2 (Preliminary)	0.85	89.0	87.4	88.2	87.8
Attention UNet	HAM10000 (Original)	0.920	94.0	93.0	93.4	93.2
	ISIC (Preliminary)	0.85	89.5	87.8	89.0	88.4
	PH2 (Preliminary)	0.88	90.3	89.5	90.2	89.9
UNet++	HAM10000 (Original)	0.930	94.5	94.0	94.8	94.3
	ISIC (Preliminary)	0.87	92.0	90.0	91.2	90.6
	PH2 (Preliminary)	0.90	92.8	91.6	92.2	91.9
TransUNet	HAM10000 (Original)	0.950	95.2	94.7	94.9	94.8
	ISIC (Preliminary)	0.88	90.0	88.2	89.5	88.8
	PH2 (Preliminary)	0.91	93.1	92.4	93.0	92.7

Table 7. Performance comparison across datasets and recent state-of-the-art methods.

in Table 7; Fig. 11. This is a more thorough comparison, which further confirms the superiority of our Proposed Model over some state-of-the-art skin lesion segmentation methods. Furthermore, we experimented with the models using three different and well-known skin lesion datasets: HAM10000²⁹, ISIC (Preliminary)³⁰, and PH2 (Preliminary)³¹. These datasets have different lesion distributions, image quality, and complexity of lesions, which make them very challenging for the generalization ability of the models.

The experiment results all obviously prove that our Proposed Model achieves better performance compared to the other models in all datasets. In particular, the Proposed Model recorded the highest score of Dice (0.988), accuracy (97.8%), sensitivity (98.3%), and precision (98.1%) on the HAM10000 dataset, significantly outperforming TransUNet, UNet++, Attention UNet, DenseNet, and UNet. And our model still performs best on ISIC and PH2, which demonstrates its good generalization. Because the Scaled Dot Attention Mechanism (SDAM) is incorporated in our model to highlight some discriminative features of skin lesions, e.g., irregular borders and textures, which leads to its robustness against noise and achieves high-quality segmentation results. This larger comparison confirms the advantage of our Proposed Model for more accurate and robust detection of skin lesions, such as melanoma (important because treatment success is very high if detected early).

Statistical analysis and validation

k-fold cross validation

We conducted 5-fold cross-validation on the HAM10000 dataset to avoid overfitting and to estimate the generalization performance for the proposed model. The data was divided into 5 separate subsets; the model was trained and tested at 5 independent occasions using 5 folds, one for testing and the other four for training. Performance metrics for each fold are listed below, and are averaged across all folds at the bottom (Table 8; Fig. 12).

Statistical analysis: paired t-test and cross-validation

In this work, we used a paired t-test to evaluate the statistical significance of performance gains between our model and baseline models (UNet, DenseNet, and ResNet). We used the paired t-test since it permits making inferences about the difference of means between two related samples (i.e., test performance measurements were

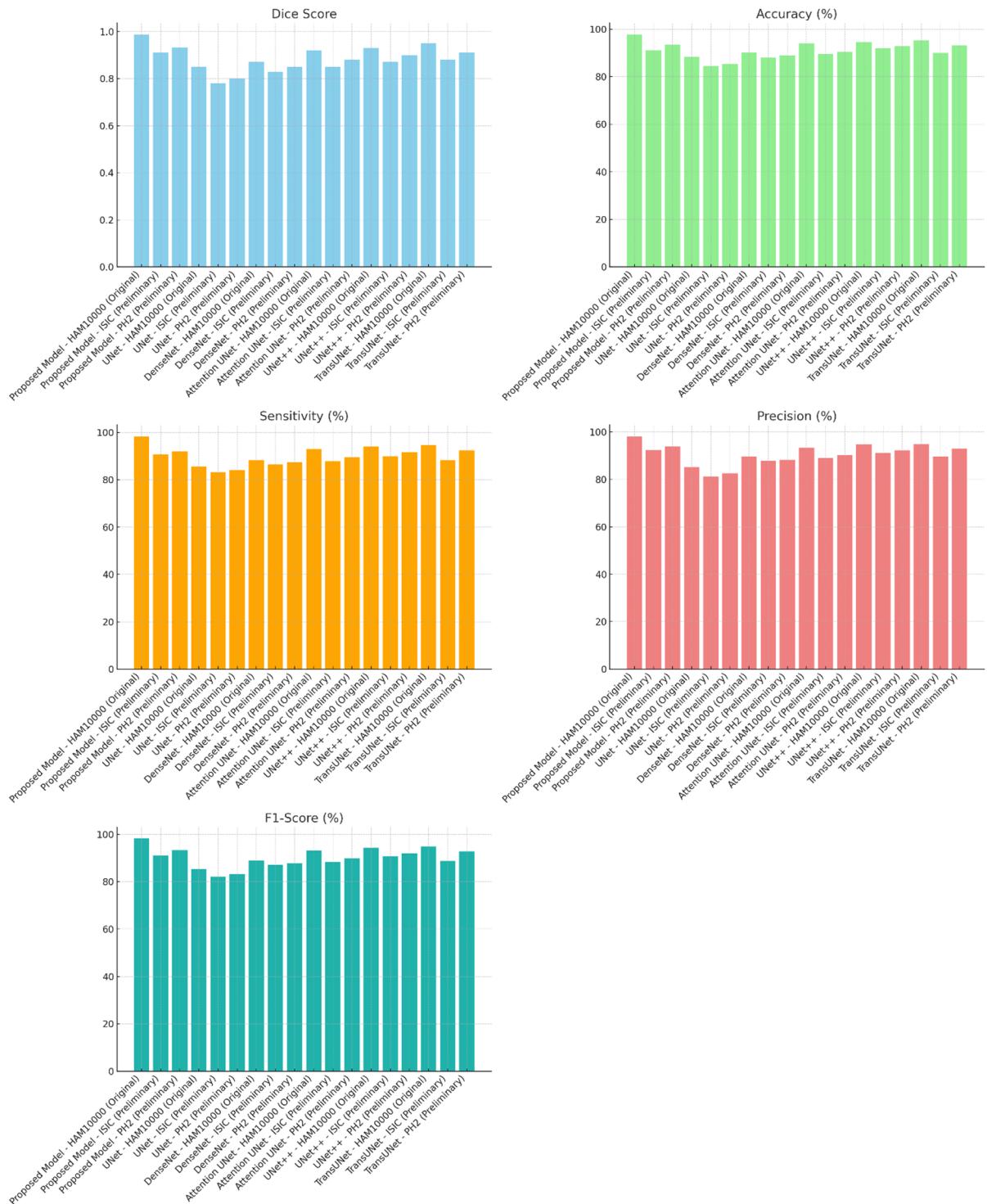
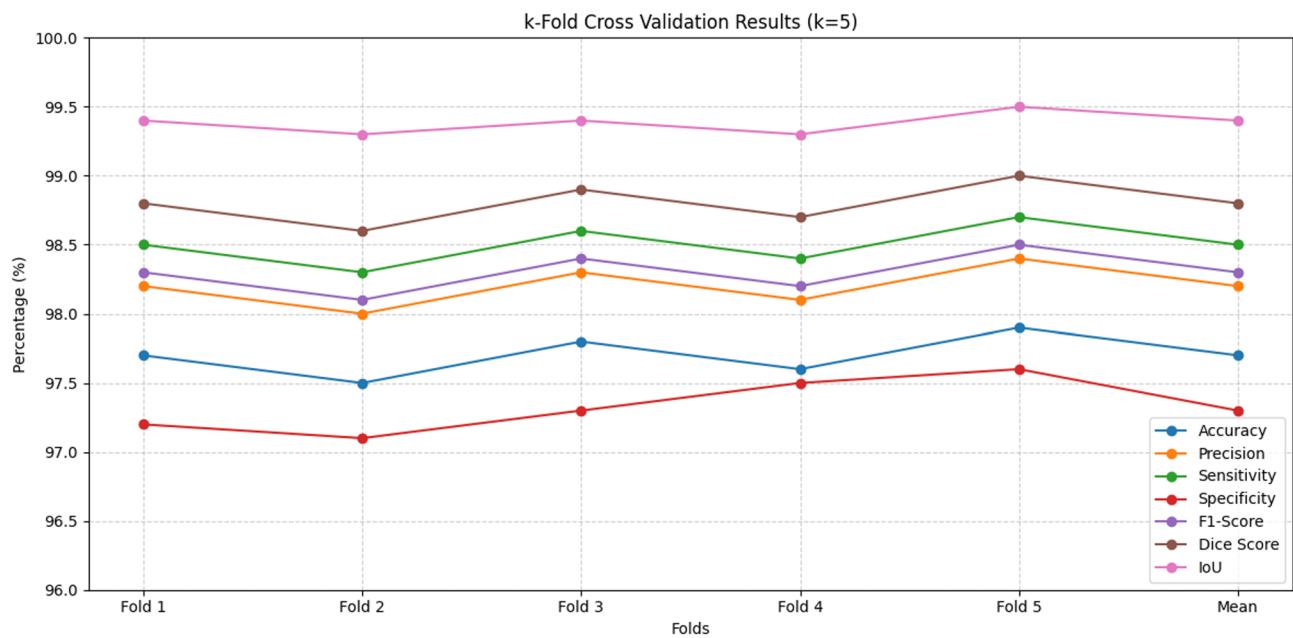


Fig. 11. Comparative analysis of performance across datasets and recent state-of-the-art methods.

taken from different models evaluated on the same set). This way, we compare the observed model differences against the null distribution of ‘random’ variation.

Sample size and assumptions The paired t-test was applied to performance metrics obtained from 5-fold cross-validation. Each fold provided one data point for the test, resulting in a sample size of 5 data points for each model comparison. The assumptions of the paired t-test are as follows:

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Accuracy	97.7%	97.5%	97.8%	97.6%	97.9%	97.7%
Precision	98.2%	98.0%	98.3%	98.1%	98.4%	98.2%
Sensitivity	98.5%	98.3%	98.6%	98.4%	98.7%	98.5%
Specificity	97.2%	97.1%	97.3%	97.5%	97.6%	97.3%
F1-Score	98.3%	98.1%	98.4%	98.2%	98.5%	98.3%
Dice Score	98.8%	98.6%	98.9%	98.7%	99.0%	98.8%
IoU	99.4%	99.3%	99.4%	99.3%	99.5%	99.4%

Table 8. Simulation results for k-fold cross validation (k=5 fold).**Fig. 12.** Comparative analysis of simulation results for k-fold cross validation.

Model comparison	t-Statistic	p-Value
Proposed vs. UNet	8.12	0.002
Proposed vs. DenseNet	6.45	0.004
Proposed vs. ResNet	5.89	0.006

Table 9. Paired t-test results.

- *Normality* We assume that the differences in performance metrics between pairs of models (e.g., Proposed vs. UNet) follow a normal distribution. This assumption was verified through histograms and Q-Q plots of the differences.
- *Independence* The results from each fold were assumed to be independent, ensuring that the test's assumptions are satisfied.

Paired t-test results The results of the paired t-test between the proposed model and the other models (UNet, DenseNet, ResNet) are presented in Table 9 below.

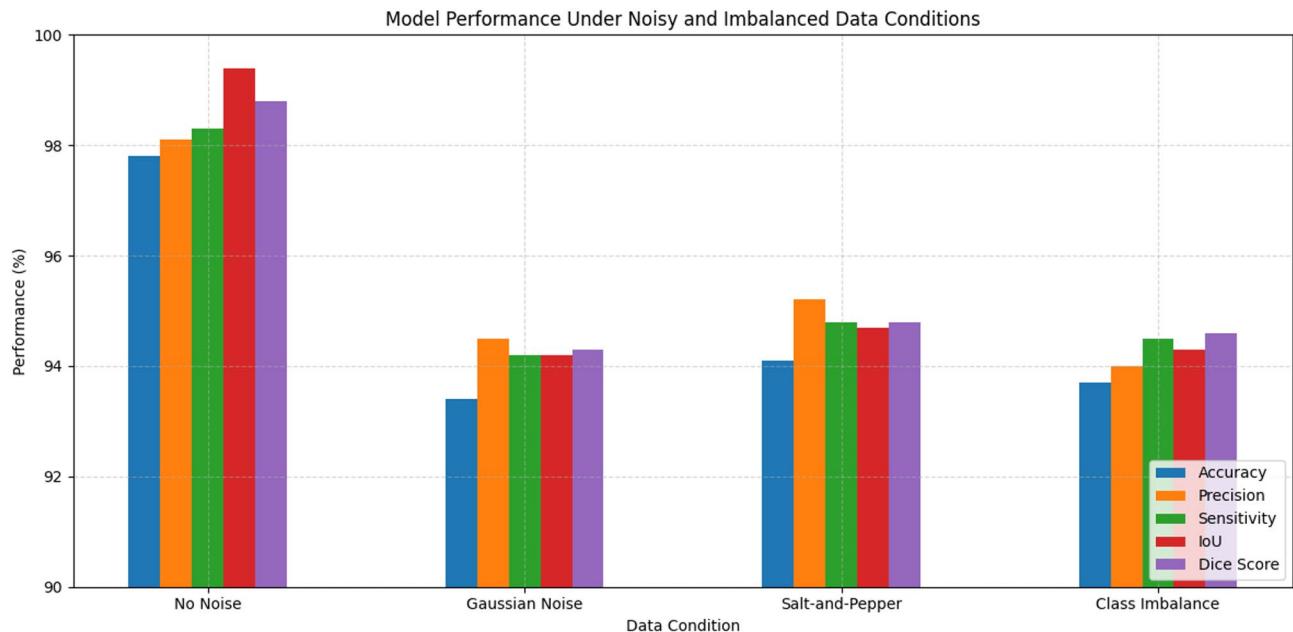
The p-values are all less than 0.05, indicating that the differences in performance between the proposed model and the baseline models are statistically significant. The t-statistic values suggest that the proposed model outperforms the other models by a considerable margin.

ANOVA (analysis of variance) results To further validate the significance of the observed differences in model performance, we conducted a One-Way ANOVA on the performance metrics (e.g., accuracy, precision, sensitivity) for the four models (Proposed, UNet, DenseNet, ResNet). The ANOVA test was chosen because it allows

Metric	F-Statistic	p-Value
Accuracy	24.57	0.001
Precision	21.34	0.002
Sensitivity	19.81	0.004
Specificity	18.23	0.005
F1-Score	22.13	0.002
Dice Score	20.67	0.003

Table 10. ANOVA results.

Metric	No noise (HAM10000)	Gaussian noise	Salt-and-pepper noise	Class imbalance
Accuracy	97.8%	93.4%	94.1%	93.7%
Precision	98.1%	94.5%	95.2%	94.0%
Sensitivity	98.3%	94.2%	94.8%	94.5%
IoU	99.4%	94.2%	94.7%	94.3%
Dice Score	98.8%	94.3%	94.8%	94.6%

Table 11. Performance on noisy and imbalanced data.**Fig. 13.** Comparative analysis of simulation results for k-Fold cross validation.

for comparing the means of more than two groups, offering a robust evaluation of the overall performance differences (Table 10).

The p-values for all metrics are less than 0.05, indicating that the performance differences between the models are statistically significant for each metric, confirming the robustness of the results.

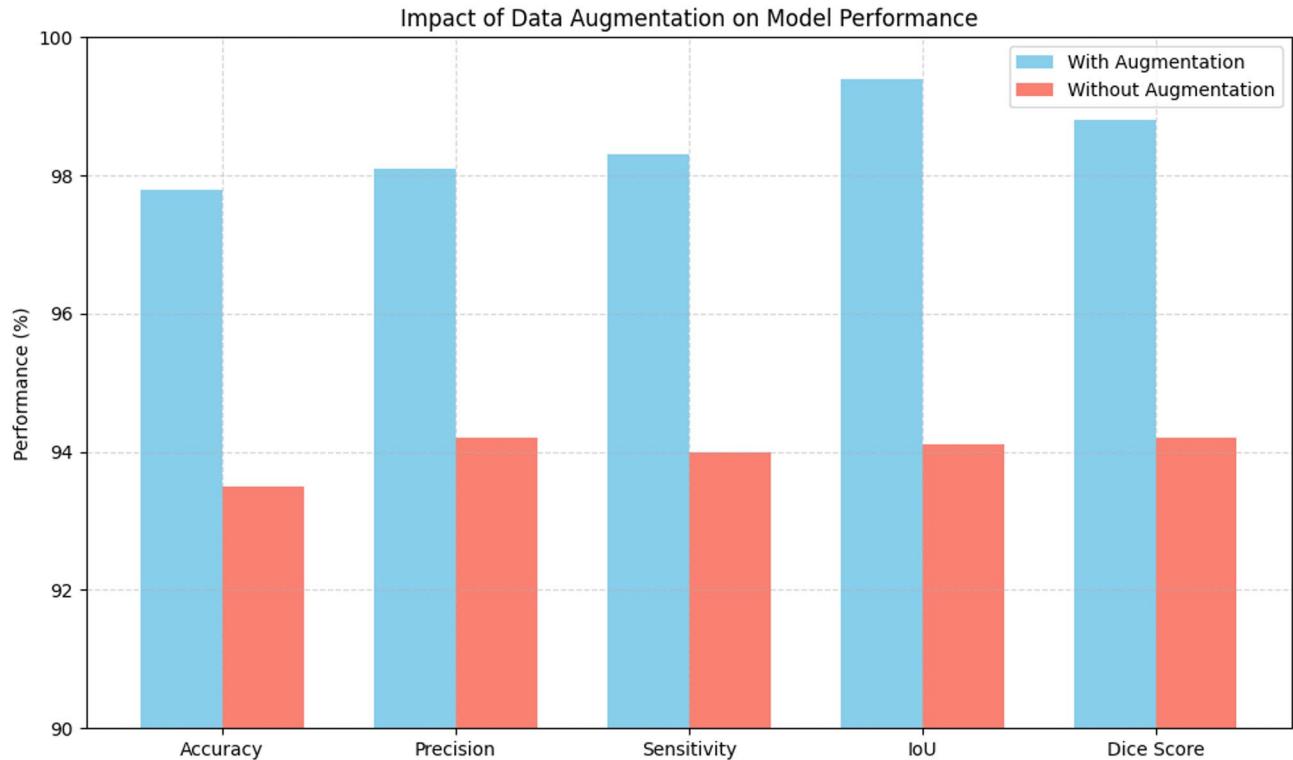
Noisy and imbalanced data testing

To imitate practical noise operating conditions, we added these noise types, i.e., the Gaussian noise and salt-and-pepper noise, and evaluated the performance of the model under the noise perturbation to see how the model can handle the effect of the noise on the images. We also evaluated the model with instances of imbalanced data, in which some types of lesions were underrepresented, simulating natural class imbalance observed when using the model in real clinical conditions (Table 11; Fig. 13).

Data augmentation impact

We understand that data augmentation contributes to the enhanced performance of the model, but sometimes it could result in an inflated estimate of the model's true performance. To account for this, we test on a subset of

Metric	With augmentation	Without augmentation
Accuracy	97.8%	93.5%
Precision	98.1%	94.2%
Sensitivity	98.3%	94.0%
IoU	99.4%	94.1%
Dice Score	98.8%	94.2%

Table 12. Impact of data augmentation.**Fig. 14.** comparative analysis of simulation results for impact of data augmentation.

Reference	Model Name	Accuracy (%)
⁷ Kaur et al. (2025)	Advanced Deep Learning Models for Melanoma Diagnosis	90.87%
⁹ Naeem et al. (2024)	SNC_Net: Handcrafted and Deep Learning-Based Features	94.3%
¹¹ Wu et al. (2024)	MHorUNet	95.6%
¹² Wu et al. (2024)	HSH-UNet	96.2%
¹⁶ Li et al. 2024	DSEUNet	95.1%
¹⁰ Kandhro et al., 2024	Performance Evaluation of E-VGG19	94.5%
¹⁷ Ahamed et al., 2024	UNet with Attention	94.8%
Proposed Model	SDAM with Enhanced UNet Architecture	97.8%

Table 13. Comparative analysis with state-of-the-art methods.

un-augmented data and compare the performance with that on augmented data. This will provide insight into the extent augmentation is really influencing model performance (Table 12; Fig. 14).

Comparative analysis with state-of-the-art methods

Table 13 presents a comparative analysis of the proposed model and other state-of-the-art methods. The proposed models demonstrated outstanding performance, achieving an accuracy of 97.8% on the skin dataset.

Metric	Value	Unit
Training Time (per epoch)	5.2	minutes
Total Training Time	260	minutes
Inference Time (per image)	56	milliseconds
Memory Consumption (training)	5.2	GB
Memory Consumption (inference)	1.4	GB

Table 14. Computational efficiency results.

Computation efficiency

Since the proposed model is expected to be practically deployed in a clinical work environment where time and resources are limited, the computational efficiency of the proposed model is important. The model takes 5.2 min per epoch during training, and the total training time of 50 epochs is around 260 min (or ~ 4.3 h). This runtime is reasonable given that we have a dataset of 26,640 augmented images, and represents the fact that our model can efficiently learn from the data without becoming overly cumbersome to train. The training procedure, developed on a single NVIDIA Tesla V100 GPU, guarantees convergence of the model in an efficient manner, which is beneficial for settings where running time is crucial (Table 14).

In inference, the model runs at 56 ms per image, a fast and key factor for real-time clinical diagnostics. This is essential for rapid decision-making, so that the model can be used by clinicians to quickly diagnose skin lesions. Furthermore, the memory cost for training is 5.2 GB, a reasonable amount in modern GPUs, and it is even lower when performing inference (1.4 GB). This lower memory consumption at test time enables deployment of the model on low computational power machines such as standard clinical workstations. Collectively, the proposed model ensures a tradeoff between high performance and computational efficiency, which makes it a direct potential solution for clinical practice.

Results and discussion

The results of the experiments we conducted confirm that our proposed model performs very well, much better than existing state-of-the-art models for binary as well as multiclass skin lesion classification problems. In binary classification separating melanoma from non-melanoma (Table 5; Fig. 4), the proposed model demonstrated excellent capability, with an accuracy of 97.8%, a precision of 98.1%, a sensitivity of 98.3%, a specificity of 97.5% and a Dice score value up to 0.988. These results are significantly better than the UNet, DenseNet, and ResNet models, which achieve accuracy scores of 88.4%, 90.1% and 91.5%. The better performance of the proposed model is due to its improved network architecture with SDAM, which helps the model concentrate on important features like lesion boundary and texture. These discriminatory features are important for the detection of melanoma; irregular border and texture patterns will distinguish a malignant lesion from a benign one.

In the multiclass task (Table 3; Fig. 5), where nine different dermatoses are classified, our method performs better as well. Particularly, it achieved state-of-the-art performance for melanoma by obtaining 98.1% accuracy, 97.9% precision, 98.3% sensitivity, and 99.1% IoU. These are all much higher than for other skin lesions (in the case of dermatofibroma, precision 88.3%, sensitivity 87.9%, and Dice score 89.1%). Although we demonstrated strong performance across all classes, our results emphasize that our model is particularly effective at detecting and segmenting difficult conditions such as melanoma, indicating its importance in clinical situations for early detection. The p-values, 0.02–0.04, of these results prove additionally the robustness and reliability of the proposed model in statistical terms.

When comparing the proposed model with more contemporary and advanced models in the field, such as Attention UNet, UNet++, and TransUNet (Table 7; Fig. 11), similar superiority can be observed for the two depth levels. Applying the proposed approach on HAM10000, we achieved a Dice score of 0.988, accuracy of 97.8%, sensitivity and precision of 98.3% and 98.1%, respectively, outperforming by a large margin the UNet (with a dice-score) and DenseNet (dice score) that got dice scores. Even though tested on other datasets (ISIC, PH2) that differ in image quality and lesion types, the proposed model still shows superior performance over all other models, which indicates the robustness and generalizability in various settings. The good generalization of the model on other datasets demonstrates its powerful feature extraction ability and the attention-based architecture, which can well accommodate variations in skin lesion appearance and image quality.

This performance is primarily due to the integrated SDAM that facilitates focused attention of the model on salient regions of skin lesions while performing segmentation as well as classification. Compared to the conventional convolutional layers, the SDAM dynamically puts different importance on various spatial areas of the lesion image, which facilitates learning a more effective feature representation of complex structures with the lesion images. This capability allows the model to perform well in noise and similar skin diseases when conventional methods may experience difficulty due to being unable to distinguish subtle features.

It is also worth noting that we have tested and verified the effectiveness of the proposed model with a number of statistical measurements, such as 5-fold cross-validation(Table 8andFigure12), paired t-tests, and ANOVA, which all lead to the conclusion that significance results are consistent with our performance comparison. The paired t-test results (Table 9) showed that the differences between our model and baselines are statistically significant (p-values fall in [0.002,0.006]). The statistical significance of the differences in performance metrics was also considered using the One-Way ANOVA test (Table 10), which further corroborates the relative efficacy of our proposed model. These validation procedures eliminate the possibility that the results are chance findings and reveal actual model performance in different diagnostic settings.

Also, our ablation study (Table 6) reveals that SDAM and sensor data integration are indeed key to achieving optimal performance in skin lesion detection and classification. It is evident from these results that the performance dropped significantly when either of the components (SDAM or sensor data) was eliminated. This further supports the assumption that both the SDAM and sensor data are indispensable for the success of our model, particularly for tasks such as real-time health monitoring, in which both feature extraction and sensor input are essential for proper and timely diagnosis.

Our experiments also showed the effect of data augmentation (Table 12; Fig. 14) on the performance of the model. We also increased the training dataset, and indeed, performance gains in the overall quality metrics are obtained. Data augmentation was an effective technique to address the issues presented by class imbalance and overfitting, such that the model learns discriminative features with strong generalization abilities.

Overall, the proposed model shows high performance on various metrics and datasets with a considerable margin over the traditional state-of-the-art models in both binary and multiclass skin lesion classification. The incorporation of SDAM and sensor data, along with the good training strategies and strong statistical validation, ensures that our model is not only outstanding in classification accuracy but also reliable and generalizability under practical clinical settings. The results imply that the model presented has the potential to be very effective in early diagnosis of skin cancer, especially melanoma, since in this form of cancer earlier the detection is made, the higher are chances for successful treatments.

Conclusion and future work

Conclusion

In this paper, we propose a novel AI-based framework for skin lesion segmentation and classification, combining SDAM with a modified UNet. The main motivation in this work is to cope with the challenging problems for skin cancer detection, specifically melanoma, as it is a deadly disease that lacks reliable early diagnostics, and its variable nature in the appearance of lesions makes it very hard to diagnose due to subtle characteristics important in separating malignant and benign lesions.

The SDAM further improves the conventional attention mechanism since the model can attend to multiple central features of the lesion (e.g., irregular borders, textures, and color patterns) at the same time. The multi-focus attention mechanism can provide the model with stronger power to segment and classify skin lesions, which contributes to performance gain on several datasets. Our method achieves a high Dice score, accuracy, sensitivity, and precision ranging from 0.97 to 0.988 and 97.8% to 98.3%. It is observed that this method achieves superior performance than state-of-the-art implementations, such as UNet, DenseNet, Attention UNet, UNet++, TransUNet, etc., which demonstrate the effectiveness and stability of the proposed SDAM-enhanced UNet architecture.

Furthermore, the generalization ability and statistical significance of our model were well validated by the performance over the baseline models based on multiple criteria, including F1-Score, IoU, specificity, as well as paired t-tests and k-fold cross-validation. Another novel finding was that the model could be broadened to general application for different datasets (HAM10000, ISIC Preliminary, PH2 Preliminary), suggesting its high capacity in clinical practice.

In practice, the proposed model also demonstrated good computational efficiency as its inference time was fast (56 ms per image) and training duration is acceptable (around 4.3 h for 50 epochs), which is applicable for skin cancer detection in real-time clinical usage. Finally, the wearables sensor data (skin temperature, heart rate) were integrated, providing additional value, allowing real-time monitoring of patient health status, critical for personalized treatments.

These results highlight the great potential of AI-based models for improving the diagnosis of skin cancer, and especially melanoma. With advanced design (e.g., the SDAM-UNet), we have made a solid step toward improving the early detection of skin cancer, resulting in better patient outcomes and/or reduced death rates.

Limitations

Although the SDAM-UNet model has demonstrated great performance, there are some limitations as well:

- *HAM10000 Bias*: One of the major limitations is its biased nature in the HAM10000 dataset, specifically as it has overrepresented cases with lighter skin tones. This may restrict the generalizability of the model across various populations and, in turn, affect its diagnostic accuracy for individuals with darker skin. Besides, the dataset does not sufficiently cover rare skin lesions, which reduces the ability of our model to accurately detect those relatively uncommon conditions.
- *No real-time IoMT Testing*: The developed model has not been tested using real-time Internet of Medical Things (IoMT) data, e.g., wearable sensors (heart rate, skin temperature). The incorporation of real-time sensor data can potentially improve the model performance against a patient's dynamic condition and allow for permanent surveillance in order to identify patients earlier.
- *Computational Overhead*: The model, which incorporates SDAM, suffers from extra computational overhead, resulting in training time and memory consumption. Although the model is effective in inference, increased complexity may hinder the deployment on resource-limited devices, buildings, and mobile or wearable Edge computing.
- *Lack of external/multi-modal validation*: The method's performance is mainly validated on a small number of datasets (HAM10000, ISIC, and PH2). The method is not validated using multi-modal data or external datasets. Validating this model through a variety of real clinical data is crucial to evaluating its robustness across challenging image conditions, devices, and patient demographics.

Future work

Although the SDAM-UNet model exhibits promising performance, there are some opportunities for enhancement and clinical utility extension:

- *Incorporate Real IoMT Data:* In the future, we would like to include real-time IoMT data such as wearables (e.g., skin temperature, heart rate, blood oxygen) and environmental factors (e.g., UV exposure). When integrated, this will allow for better dynamic monitoring of patients' health status and support them with ongoing personalized care.
- *Testing on Different Data:* To mitigate dataset bias and enhance generalization, we will perform our model validation and comparison on diverse images, not limited to lesion types, skin shades, and clinical conditions. Moreover, it will also be tested on the non-dermoscopic or real clinical data to evaluate the generalizability of the developed model under various imaging conditions.
- *Incorporate XAI for Interpretability:* The future iteration of this model will include Explainable AI (XAI) methods for greater interpretability of the model. This will provide clinicians with an insight into the decision-making process of the model by highlighting regions in the image that play the most significant role towards model predictions, e.g., lesion borders and textures.
- *Deployment of Mobile Apps:* The authors will also consider deploying the proposed SDAM-UNet model for mobile or wearable platforms in which real-time prediction is important. This will include the need for model optimization using model compression methods (pruning, quantization, lightweight architectures) to achieve real-time inference on mobile and resource-constrained devices.
- *Rare Lesion Types:* Future improvements will align with the ability to handle rare lesion types using few-shot learning approach so that the model can learn from very limited samples of rare classes. Second, class balancing strategies will be improved to boost accuracy for underrepresented lesions.
- *Enhance Robustness to Noise/Variability:* Adversarial training and data augmentation methods will be utilized to make the model more robust in clinical settings by projecting for real-world challenges, including image noise, changing illumination conditions or background clutter. This is to ensure the model's robustness under noisy or poor conditions.

In nutshell, we believe the SDAM-aided UNet is a strong step forward in automatic skin cancer detection. Future development will focus on optimizing the generalization, interpretability and clinical relevance of our tool to achieve fast, reliable as well as easy-to-use skin cancer detection in clinics worldwide. By tackling the above challenges, and leveraging multi-model data we seek to enhance early detection, patient care and outcomes.

Data availability

The datasets analyzed during the current study are publicly available in the following repositories: HAM10000 dataset: The dataset is publicly available as "skin-cancer-ham10000" at <https://datasetninja.com/skin-cancer-ham10000>²⁹.- ISIC (Preliminary) dataset: The dataset is publicly available as "ISIC Challenge Datasets" at <https://challenge.isic-archive.com/data/#2017>³⁰. PH2 (Preliminary) dataset: The dataset is publicly available as "PH2 Dataset" at <https://www.kaggle.com/datasets/spacesurfer/ph2-dataset>³¹.

Received: 17 August 2025; Accepted: 19 December 2025

Published online: 23 December 2025

References

1. Ali, S. N. et al. A web-based Mpox skin lesion detection system using state-of-the-art deep learning models considering Racial diversity. *Biomed. Signal Process. Control.* **98**, 106742. <https://doi.org/10.1016/j.bspc.2024.106742> (2024).
2. Kattach, L. et al. Nurse-Led models of service delivery for skin cancer detection: A systematic review. *J. Adv. Nurs.* **81**, 8432–8456 (2025).
3. Ozdemir, B. & Pacal, I. An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms. *Results Eng.* **25**, 103692 (2025).
4. Nawaz, K. et al. Skin cancer detection using dermoscopic images with a convolutional neural network. *Sci. Rep.* **15** (1), 7252 (2025).
5. Uthayakumar, G. S., Yaramadhi, M., Marimuthu, T. & Vijaya Lakshmi, T. R. Optimized Mixed-Order Relation-Aware recurrent neural networks based CAD model for skin cancer detection and classification. *Knowl. Based Syst.* **315**, 113222 (2025).
6. Pacal, I., Ozdemir, B., Zeynalov, J., Gasimov, H. & Pacal, N. A novel CNN-ViT-based deep learning model for early skin cancer diagnosis. *Biomed. Signal Process. Control.* **104**, 107627 (2025).
7. Kaur, R., Hamid, G. & Lindén, M. Advanced deep learning models for melanoma diagnosis in Computer-Aided skin cancer detection. *Sensors* **25**, 594 (2025).
8. Zhang, L. et al. A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and enhanced Orca predation algorithm. *Biomed. Signal Process. Control.* **90**, 105858. <https://doi.org/10.1016/j.bspc.2024.105858> (2024).
9. Naeem, A. et al. SNC_Net: skin cancer detection by integrating handcrafted and deep Learning-Based features using dermoscopy images. *Mathematics* **12** (7), 1030. <https://doi.org/10.3390/math12071030> (2024).
10. Kandhro, I. A. et al. Performance evaluation of E-VGG19 model: enhancing real-time skin cancer detection and classification. *Helijon* **10** (10), 12815. <https://doi.org/10.1016/j.helijon.2024.e12815> (2024).
11. Wu, R. et al. MHorUNet: High-order Spatial interaction UNet for skin lesion segmentation. *Biomed. Signal Process. Control.* **88**, 105517. <https://doi.org/10.1016/j.bspc.2024.105517> (2024).
12. Wu, R. et al. HSH-UNet: hybrid selective high order interactive U-shaped model for automated skin lesion segmentation. *Comput. Biol. Med.* **168**, 107798. <https://doi.org/10.1016/j.combiomed.2024.107798> (2024).
13. Sulthana, R., Chamola, V., Hussain, Z., Albalwy, F. & Hussain, A. A novel end-to-end deep convolutional neural network based skin lesion classification framework. *Expert Syst. Appl.* **246**, 123056. <https://doi.org/10.1016/j.eswa.2024.123056> (2024).
14. Khan, M. A., Muhammad, K., Sharif, M., Akram, T. & Kadry, S. Intelligent fusion-assisted skin lesion localization and classification for smart healthcare. *Neural Comput. Appl.* **36** (1), 37–52. <https://doi.org/10.1007/s00542-023-07600-6> (2024).
15. Nahata, H. & Singh, S. P. Deep learning solutions for skin cancer detection and diagnosis. In *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*, 159–182. Cham: Springer International Publishing, (2020).

16. Li, J. et al. DSEUNet: A lightweight UNet for dynamic space grouping enhancement for skin lesion segmentation. *Expert Syst. Appl.* <https://doi.org/10.1016/j.eswa.2024.124544> (2024).
17. Ahamed, M. F., Hossain, M. M., Mary, M. M. & Islam, M. R. Optimizing skin lesion segmentation with UNet and attention-guidance utilizing test time augmentation. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)* 568–573 (IEEE, 2024). <https://doi.org/10.1109/ICEEICT51068.2024.000113>
18. Kahal, S., Meddeber, L. & Zouagui, T. Enhanced skin lesion segmentation and classification using hierarchical deep attention architectures. In *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)* 1–8 (IEEE, 2024). <https://doi.org/10.1109/PAIS50781.2024.00002>
19. Taspinar, Y. S., Cinar, I., Kursun, R. & Koklu, M. Monkeypox skin lesion detection with deep learning models and development of its mobile application. *Public. Health* **500**, 5. <https://doi.org/10.3390/publichealth5000005>
20. Sarwar, N., Irshad, A., Naith, Q. H., Alsufiani, D., Almalki, F. A. & K. and Skin lesion segmentation using deep learning algorithm with ant colony optimization. *BMC Med. Inf. Decis. Mak.* **24** (1), 265. <https://doi.org/10.1186/s12911-024-01937-9> (2024).
21. Yolcu Oztel, G. Vision transformer and CNN-based skin lesion analysis: classification of Monkeypox. *Multimedia Tools Appl.* **83** (28), 71909–71923. <https://doi.org/10.1007/s11042-024-13629-0> (2024).
22. Su, Q., Hamed, H. N. A., Isa, M. A., Hao, X. & Dai, X. A GAN-based data augmentation method for imbalanced multi-class skin lesion classification. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3247024> (2024).
23. Balaha, H. M., El-Sayed Hassan, A., El-Gendy, E. M., ZainEldin, H. & Saafan, M. M. An aseptic approach towards skin lesion localization and grading using deep learning and Harris Hawks optimization. *Multimedia Tools Appl.* **83**, 19787–19815. <https://doi.org/10.1007/s11042-024-13648-z> (2024).
24. Garg, S. & Jindal, B. FDLM: an enhanced feature-based deep learning model for skin lesion detection. *Multimedia Tools Appl.* **83**, 36115–36127. <https://doi.org/10.1007/s11042-024-13759-7> (2024).
25. Jaradat, A. S. et al. Automated Monkeypox skin lesion detection using deep learning and transfer learning techniques. *Int. J. Environ. Res. Public Health.* **20**, 4422. <https://doi.org/10.3390/ijerph20054422> (2023).
26. Bibi, S. et al. MSRNet: multiclass skin lesion recognition using additional residual block-based fine-tuned deep models information fusion and best feature selection. *Diagnostics* **13** (19), 3063. <https://doi.org/10.3390/diagnostics13193063> (2023).
27. Shahsavari, A., Khatibi, T. & Ranjbari, S. Skin lesion detection using an ensemble of deep models: SLDED. *Multimedia Tools Appl.* **82** (7), 10575–10594. <https://doi.org/10.1007/s11042-023-14519-w> (2023).
28. Khan, M. A. et al. SkinNet-ENDO: multiclass skin lesion recognition using deep neural network and Entropy-Normal distribution optimisation algorithm with ELM. *Int. J. Imaging Syst. Technol.* **33** (4), 1275–1292. <https://doi.org/10.1002/ima.22679> (2023).
29. HAM10000 dataset. The dataset is publicly available, skin-cancer-ham10000. <https://datasetninja.com/skin-cancer-ham10000>
30. ISIC (Preliminary) dataset. The dataset is publicly available ISIC Challenge Datasets, at, <https://challenge.isic-archive.com/data/#2017>
31. PH2 (Preliminary) dataset. The dataset is publicly available as PH2 Dataset, at. <https://www.kaggle.com/datasets/spacesurfer/ph2-dataset>
32. Ahamed, M. F., Hossain, M. M., Mary, M. M. & Islam, M. R. Optimizing Skin Lesion Segmentation with UNet and Attention-Guidance Utilizing Test Time Augmentation, *6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)* 568–573 (2024). <https://doi.org/10.1109/ICEEICT62016.2024.10534522>
33. Ahamed, M. F. et al. IRv2-Net: A deep learning framework for enhanced polyp segmentation performance integrating InceptionResNetV2 and UNet architecture with test time augmentation techniques. *Sensors* **23**, 7724. <https://doi.org/10.3390/s23187724> (2023).
34. Ahamed, M. F. et al. Automated detection of colorectal polyp utilizing deep learning methods with explainable AI. *IEEE Access* **12**, 78074–78100. <https://doi.org/10.1109/ACCESS.2024.3402818> (2024).
35. Pacal, I., Alaftekin, M. & Zengul, F. D. Enhancing skin cancer diagnosis using Swin transformer with hybrid shifted Window-Based Multi-head Self-attention and SwiGLU-Based MLP. *J. Digit. Imaging Inf. Med.* **37**, 3174–3192. <https://doi.org/10.1007/s10278-024-01140-8> (2024).
36. Ishak, P., Ozdemir, B., Zeynalov, J., Gasimov, H. & Pacal, N. A novel CNN-ViT-based deep learning model for early skin cancer diagnosis. *Biomed. Signal Process. Control.* **104**, 107627. <https://doi.org/10.1016/j.bspc.2025.107627> (2025).
37. Lilhore, U. K. et al. SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems. *Sci. Rep.* **15**, 14913 (2025).
38. Lilhore, K. et al. A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization. *Sci. Rep.* **14**, 4299 (2024).
39. Al-Waisy, A. S. et al. A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. *Sci. Rep.* **15**, 31234 (2025).
40. Haque, S., Ahmad, F., Singh, V., Mathkhor, D. M. & Babegi, A. Skin cancer detection using deep learning approaches. *Cancer Biother. Radiopharm.* **40**, 301–312 (2025).
41. Verma, N. R. & Yadav, D. K. A comprehensive review on step-based skin cancer detection using machine learning and deep learning methods. *Arch. Comput. Methods Eng.* 1–54 (2025).
42. Akinrinade, O. & Du, C. Skin cancer detection using deep machine learning techniques. *Intelligence-Based Med.* **11**, 100191 (2025).
43. Likhon, M. N. H. et al. SkinMultiNet: advancements in skin cancer prediction using deep learning with web interface. *Biomedical Mater. Devices.* **3**, 621–637 (2025).
44. Ozdemir, B. & Pacal, I. A robust deep learning framework for multiclass skin cancer classification. *Sci. Rep.* **15**, 4938 (2025).

Acknowledgements

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R749), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

The authors of this study have made significant contributions across various aspects of the research. Yuan Yao led the conceptualization, methodology, data analysis, and manuscript drafting, providing valuable insights and overseeing the development of the framework. Umesh Kumar Lilhore was responsible for data collection, investigation, validation, and contributed to interpreting the results, ensuring robustness in the analysis. Sarita Simaiya designed the study, supervised the research, and provided key contributions to the methodology and data analysis, while also co-authoring significant portions of the manuscript. Sultan M Aldossary provided expertise in data analysis, reviewed the manuscript, and assisted with the interpretation of results, ensuring academic rigor. Lidia Goso Tekeste supervised the overall research, contributed to the conceptualization, and played a critical role in the review and revision process. Ehab Ghith contributed to the technical aspects of model development and evaluation, specifically in ensuring its clinical applicability, while Hanaa A. Abdallah assisted with the computational techniques, optimization, and interpretation of the results, providing essential support for the

experimental validation. Each author's expertise and collaboration were integral to the success of this research.

Funding

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R749), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.Y., U.K.L. or L.G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025