RESEARCH PAPER

An Attention-Guided Dual-Branch Deep Learning Framework for Accurate Segmentation and Classification of Skin Lesions in Melanoma Detection

Charanya Vanditha[1*], Raghavendra Reddy[2]

[1*] School of Computer Science and Application, REVA University, Bangalore, India. (Corresponding Author)

Email: charanyavn@gmail.com

[2] School of Computer Science and Engineering, REVA University, Bangalore, India.

Received: 20th Feb, 2026 | **Revised:** 4th Mar, 2026 | **Accepted:** 25th Mar, 2026 | **Available Online:** 10th Apr, 2026

ABSTRACT

The accurate segmentation and classification of skin lesions are critical to detect melanoma at an early stage, but the classical machine learning algorithms are usually unreliable and the recent deep learning systems are characterized by issues of computational complexity and generalization. To overcome these shortcomings, this paper suggests a hybrid deep learning model that can be used to perform lesion segmentation and categorization simultaneously. The model includes four steps, which are preprocessing, dual-branch encoding, multi-scale decoding, and lesion classification. The preprocessing module improves the quality of images by means of contrast regulation, normalization, rescaling, and hair removal. The encoder has two branches, one branch pours out complementary based local texture and global shape features and the other pulls out the features and blends them through attention-based fusion mechanism in order to emphasize the most informative ones. With a multi-task decoder that includes dilated convolutions and skip connections, the boundaries of lesions can be refined even in low-contrast areas. Lastly, lesion type is predicted by a classification head with fully connected layers with a softmax classifier. The experimental analysis is carried on benchmark dermoscopic datasets (PH2, ISIC 2017–2020, and HAM10000) and outcome of these experiments demonstrates the effectiveness of proposed approach. The proposed approach has reported Dice Similarity Coefficient of 0.985 and corresponding Intersection Over Union (IoU) of 0.970 which indicates robust capability for lesion boundary delineation and classification.

Keywords: Skin lesion segmentation, Melanoma classification, Deep learning, Dual-branch encoder, Attention-guided fusion.

How to cite this article: Vanditha C, Reddy R. An Attention-Guided Dual-Branch Deep Learning Framework for Accurate Segmentation and Classification of Skin Lesions in Melanoma Detection. Int J Drug Deliv Technol. 2026;16(32s):83-99. DOI: 10.25258/ijddt.16.32s.10

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

1. Introduction

Skin is the largest organ of the human body, which has the function of acting as the key protective layer against external factors, thermoregulation, and sensory perception, including touch, heat, cold, and others [1]. Nevertheless, genetic mutations or exposure to the environment can interfere with the normal developmental pathway of skin cells causing multiple skin cancers. Among them, melanoma is the most aggressive and life-threatening, even though this type of cancer is less common than non-melanoma skin cancers [2]. Melanoma arises in melanocytes pigment forming cells formed as a result of neural crest neoplasia and has a great metastatic tendency unless diagnosed and treated at an early age [3 ,4]. The World Health Organization (WHO) has reported that the incidence of both melanoma and non-melanoma skin cancer in the world has increased drastically in the last several decades. The world has around 132,000 cases and 2-3 million cases of non-melanoma skin cancer, each year [5]. According to the American Cancer Society, melanoma represents only the case of about 4% of all skin cancer, but the cause of almost three-fourths of deaths caused by skin cancer [6]. There were 197,700 estimated cases of new melanoma in 2022 in the USA alone, both invasive and in situ [7]. Moreover, estimates given by the International Agency for Research on Cancer (IARC) reveal that the frequency of melanoma will go up by more than 50% by 2040, which means approximately half a million new cases and approximately one hundred thousand deaths each year [8]. Melanoma can occur in different population groups, with differences in the incidence and mortality rate being largely dependent on ethnicity, skin type, and ultraviolet (UV) exposure [9]. Melanoma is common in places of fair-skinned people and high UV radiation like Australia and New Zealand where the rates of the disease stand at 42 cases per 100,000 men and 31 per 100,000 women [10]. On the other hand, most of the African and

Asian countries have an incidence that is less than 1 case per 100,000 people. The known risk factors are long term long-term exposure to UV, living in low-latitude areas, high intake of alcohol or fatty food, phenotypic traits (fair skin, light hair, freckles), and genetic factors. The most important of them is UV radiation, which is the etiological factor causing the malignant transformation of melanocytes through the mutation of DNA [11]. When this transformation takes place abnormally, the growth of melanocytes becomes abnormal, and fibrous tissues grow in the deeper dermal layers leading to the cancerous lesions that could spread to the epidermis and then manifest themselves on the surface of the skin. Figure 1 depicts the melanoma and its spread in the skin.

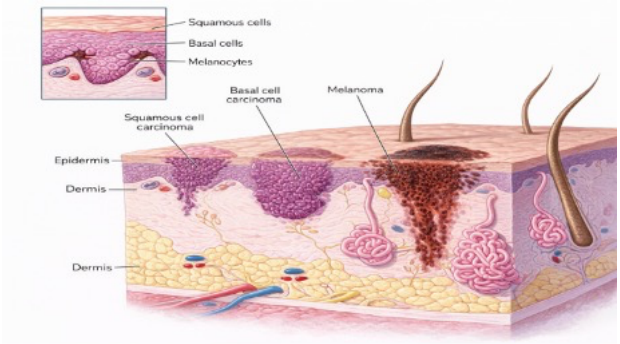

Figure 1 Melanoma and it's spread in skin

In most cases malignant lesions are similar to benign moles in size, shape and pigmentation and therefore detecting them at the early stages using the naked eye is very difficult. The five types of skin cancer as illustrated in Fig. 2 include melanoma (MEL) fig 2(i), basal cell carcinoma (BCC) fig 2(ii), benign keratosis-like lesions (BKL) fig 2(iii), congenital melanocytic nevus (CMN) fig 2(iv), and benign lesions (BEN) fig 2(v) [12]. The most aggressive and life-threatening of them is melanoma that can spread over tissues in the surrounding area and to other organs. The other forms are mostly benign and are slow growing, well demarcated, and have minimal chances of invasion. Therefore, it is necessary to distinguish and correctly identify melanoma and harmless lesions to intervene in time and increase the chance of survival.

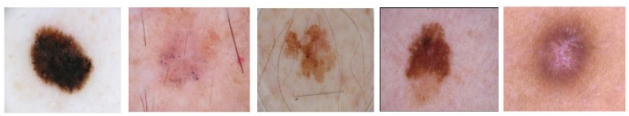

Figure 2 Images of different skin cancer types: (i) MEL, (ii) BCC, (iii) BKL, (iv) CMN, (v) BEN.

Even with the five-year survival rate being about 98% in the case of early-stage melanoma, the survival rate declines drastically when the disease reaches the advanced stages and the survival rate is only 20-50% [13]. Consequently, the early identification and accurate delineation of lesions are crucial in enhancing patient outcomes and minimizing death cases. Historically, visual inspection was the diagnostic method used in dermatology, with around 60% diagnostic accuracy, and dermoscopic imaging raised the diagnostic accuracy to 75-84%. However, the dermoscopic interpretation is still subjective, time-consuming, and has inter-observer variability and the need to develop automated, strong, and reproducible computer-assisted diagnostic (CAD) systems is urgent.

1.1 Problem statement

Melanoma is the worst of all skin cancer and causes most of the deaths associated with skin cancer though it is a minor percentage of all the skin cancers. Early diagnosis has a dramatic positive effect on survival, but due to the visual similarity of both the malignant and benign lesion, changes in the shape, color, and texture of the lesion, and because of the artifact that includes hair and low-contrast delimiting areas in the dermoscopic images, it is difficult to diagnose the lesion accurately and in a timely manner. Conventional methods of diagnosis such as naked-eye examination and dermoscopy are subjective, open to inter-observer errors, and lack accuracy. At present, the automated segmentation and classification solutions also have essential shortcomings, such as low work on non-homogenous and noisy images, reliance on manual features, lack of generalization between datasets, and prohibitive computational expenses. Therefore, the necessity is strong to have a resilient, precise, and entirely automated framework that is able to perform dependable melanoma lesion segmentation and classification under variable imaging circumstances.

1.2 Motivation

The increase in the rate of melanoma in the world, which is estimated to grow over 50 % by the year 2040, exerts more burden on the health care systems, particularly in those areas that are not equipped with trained dermatologists. Despite the incredible capabilities of deep learning in medical image analysis, the practical application of the technology to melanoma detection is still challenging because of insufficient annotated data, overfitting, varying imaging scenarios, and inaccurate segmentation. Introducing a hybrid system that combines both state-of-the-art deep learning structures and powerful preprocessing and segmentation algorithms can greatly minimize diagnostic errors. This creates the urge to have a streamlined, precise, and universally applicable automated system that can process real-world dermoscopic images with the goal of a reliable melanoma segmentation and classification.

Several researches have been carried out to automatically detect, segment and classify the melanoma lesions with the help of dermoscopic images. Melanoma segmentation plays important role but these methods face several problems because of poor image contrast, contour irregularities, color differences and existence of several artifacts such as hair or bubble and heterogeneous shapes of lesions. In order to overcome these issues, researchers have developed machine learning based approaches where data pre-processing, hand-crafted feature extraction and machine learning classifiers are used to train and classify the patterns. ML-based approaches often require handcrafted features and perform poorly on complex or noisy images. Deep learning their variants, have revolutionized melanoma segmentation by learning hierarchical and discriminative representations from large-scale image data. Despite these advances, deep learning models still face issues such as limited annotated datasets, overfitting, computational complexity, and poor generalization to unseen data. Rest of the article is organized as follows: section II presents the brief literature review, section III presents the detailed discussion about proposed approach and its various components, section IV presents the outcome of proposed approach and its comparative analysis with state-of-art methods and finally section V presents the concluding remarks about proposed work.

2. Literature review

This section presents a brief historical review of the current state-of-the-art methods in the field of skin lesion segmentation in melanoma detection and classification.

2.1 Machine learning approaches

Early research on automated melanoma segmentation were mostly based on ML and conventional image processing methods. These methods are usually carried out in following stages: preprocessing, feature extraction, segmentation and classification to differentiate between malignant melanoma and benign lesions.

According to Rehman et al. [14] discussed the major drawbacks of the current segmentation tools, which fail at handling dermoscopic data having artifacts like corner edges, low-contrast hair, and varying lesion size. To overcome this issue, authors proposed an improved segmentation pipeline that included an artifact removal phase, image enhancement, and GrabCut segmentation, which had better delineation accuracy than conventional approaches.

Similarly, Imtiaz et al. [15] developed a combined approach using image processing and ML based framework to overcome the low-contrast dermoscopic images that usually inhibit the accuracy of segmentation and classification. The pre-processing phase considers noise removal, hair removal, RGB intensity adjustments, Otsu thresholding and morphological operations to clean up the lesion masks. To overcome the overfitting issue, this model utilizes Synthetic Minority Over-Sampling Technique (SMOTE. Finally, color and ABCD characteristics are used as features and fed into the classifier to obtain the final classification performance.

Verman et al. [16] introduced a hybrid approach for segmentation and classification using CNN based approach. The first phase performs data pre-processing where noise removal is performed using Gaussian filter and later the filtered image is processed through the GrabCut algorithm for segmentation. further, the obtained segmented images were processed through the CNN architecture constructed using three hidden layers and Adam optimizer to perform the classification.

Faizi et al. [17] introduced a melanoma detection framework with template matching using normalized cross-correlation to obtain the region of interest (ROI). The histogram equalization and k-means clustering were used to enhance contrast and segment respectively. Hu moments and GLCM-based Haralick features were used by the authors to extract shape features and characterise texture respectively. Several classifiers such as the random forest, the SVM, the decision tree, and the KNN were tested, and they attained a good classification of benign and malignant lesions.

These methods emphasize the advantages of standard ML-based segmentation techniques in working with the moderate sized datasets and obtaining interpretable results. However, these approaches are extremely sensitive to image quality, light change and lesion abnormalities. Moreover, these methods rely on handcrafted characteristics, a low feature discriminability level, and hand preprocessing steps that limit their scalability and resistance in the real-world clinical environment. This has led the research community to increasingly adopt deep learning based techniques, specifically CNNs and encoder-decoder systems that are capable of automatically learning hierarchical representations of raw dermoscopic images. The following sub-section presents the developments and issues related with these deep learning methods of skin lesion segmentation and melanoma detection.

2.2 Deep learning-based methods

Behera et al. [18] highlighted the limitations of traditional encoder-decoder networks, such as DenseNet and ResNet-based designs, which often struggle with complex parameter

settings, lack of multi-scale information, and insufficient pre-trained features. Their work employed UNet and ResUNet architectures to provide high-quality lesion segmentation, demonstrating robustness across diverse dermoscopic images. ThangaPurni et al. [19] reported that tradition neural network and Transformer based methods struggle to capture the fine grained local details and long range contextual information. To overcome this issue, authors developed a hybrid ARP-ViT-CNN model which employs three stream feature extraction process by employing Angular Radial Partitioning (ARP), CNN, and Vision Transformers. The ARP helps to extract the geometric and structural features, ViT helps to code the long-range dependencies by using self-attention mechanism. A comprehensive pre-processing pipeline, including dataset balancing and enhancement, further improved the model's generalization across underrepresented classes. Kumar et al. [20] introduced a deep learning-based approach for segmentation and classification. This method presents a novel approach where GLCM feature extraction with DeepLabV3+ is used for lesion segmentation. Later, a Long Short-Term Memory (LSTM) network is trained on these features obtained by applying GLCM. This combined method was successful in capturing the differences in textures and the contextual information that is essential in the differentiation between benign and malignant lesions.

Mustafa et al. [21] presented a three-stage hybrid deep learning framework for data pre-processing, segmentation and classification. The pre-processing phase includes morphology based technique later the obtained image is processed through the segmentation phase where advanced deep learning ResUNet++ is used and later, combination of transfer learning based approach using AlexNet and Random Forest classifier were used for robust lesion classification. Albahli et al. [22] proposed real-time detection and segmentation model using YOLOv8, and the training strategy is to focus on multi-dataset training to increase the ability to generalize. The pre-processing techniques include adaptive contrast enhancement, and artefact removal were used whereas advanced augmentation methods such as CutMix and Mosaic were applied to enhance the lesion diversity. The YOLOv8 architecture presents a unified lesion detection and segmentation in a single inference pass that helps to significantly enhance the computational efficiency. Anand et al. [23] developed a fusion model combining U-Net for segmentation and CNN for multi-class classification, highlighting the importance of accurate lesion isolation for subsequent classification tasks. Similarly, Bindhu et al. [24] presented a novel MFO-Fuzzy U-Net for segmentation, where preprocessing with bilateral filters and optimization with the May Fly Optimizer enhanced accuracy. In this work, authors used an IoT-based system to capture images using a Raspberry Pi and bilateral filter to eliminate unwanted noise and artifacts.

Although these deep learning methods have been capable of performing remarkably better, there are still a couple of limitations. Most models need large annotated datasets to train, have difficulties with small or low-contrast lesions, and can overfit to particular datasets, which restricts generalizability. Moreover, the way that some of the approaches are computationally intensive makes real-time deployment difficult. Artifacts, changing light conditions, and heterogeneity of the skin make segmentation even more challenging, which encourages the creation of more efficient, lightweight, and generalizable deep learning models capable of addressing the requirements of real-world clinical settings.

Table 1. Notations used in this work

Symbol	Description
$I_{in}\ \mathbb{R}^{H\times W\times C}$	Input raw dermoscopic image
$F_t^{(l)}$	Feature map of layer (l) in the texture branch
L_t	Number of layers in the texture branch
$F_s^{(l)}$	Feature map of layer (l) in the shape branch
L_s	Number of layers in the shape branch
$*\,d$	Dilated convolution operation
F_t^{concat}	Concatenated texture features
F_{fusion}	Fused feature map
F_{CA}	Channel-attended fused feature map
D^l	Decoder feature map
$*\,u$	Unsampling convolution
$Y_{seg} \in [0,1]^{H\times W}$	Predicted segmentation mask
Y_b	Ground truth boundary map
Y_{cls}	One-hot ground truth lesion class label
$\hat{Y}_{seg}, \hat{Y}_b, \hat{Y}_c$	Predicted outputs for segmentation, boundary, and classification heads
L_{seg}	Segmentation loss: weighted sum of BCE and Dice loss

L_b	Boundary refinement loss:
L_{cls}	Classification loss
$\lambda_1, \lambda_2, \lambda_3$	Adaptive weights for loss components
$F_{concat}, f_{energy}, f_{homogenity}$	Handcrafted GLCM feature descriptors
$H(x, y)$	Binary mask for hair/artifact regions
$G(i, j; \sigma)$	2D Gaussian kernel
$\eta(x, y)$	Additive noise

3. Proposed model

The UNet architectures have been widely adopted in various segmentation tasks such as UNet, UNet++, ResUNet, ResUNet++ etc. therefore, we adopt this architecture as base architecture and introduce a Dual-Branch Texture–Shape Fusion Network (TSFNet) which is an advanced UNet-inspired multi-tasking deep learning enabled architecture which is designed to enhance the segmentation accuracy by jointly modelling the texture and shape cues. The conventional architecture such as UNet and ResUNet++ are generally focus on hierarchical spatial encoding to perform the segmentation however, these methods fail to preserve the fine-grained texture and boundary information. This affects the overall segmentation performance. To overcome this issue, the proposed TSFNet introduces two dedicated branches: texture branch and shape branch which are mainly focused on extracting the texture and shape features of given input data. The obtained features are then fused through the channel-spatial attention fusion model (CSAFM). The proposed model is able to predict the segmentation masks, boundary map and lesion class labels as unified framework. Figure 3 depicts the overall architecture of proposed model.

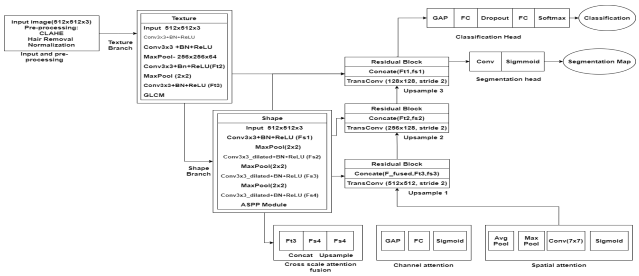

Figure 2 Proposed architecture for segmentation and classification

3.1 Data Pre-processing

This is the first stage of proposed model where dermoscopic images are processed through multiple steps. This phase produces standardized image from raw image and discards different type of irregularities to ensure smooth learning and consistent high-quality input. Let us consider that the raw image is represented

$$I_{in} \in \mathbb{R}^{H \times W \times C} \quad (1)$$

Where H, W, C represents the height, width and channel of image respectively. The outcome of final pre-processed image is represented as:

$$I_p \in \mathbb{R}^{H_t \times W_t \times C} \quad (2)$$

The pre-processed image generates normalized image which includes resizing, anti-alising, and intensity normalization and standardization and contrast enhancement. Further, image denoising and artefact removal such as hair, ruler marks, and bubbles is applied to enhance the quality of image. Structural information. The bicubic interpolation method provides superior smoothness and preserves the information detail. This approach estimates the intensity value $I_p(x, y)$ as resampled pixel at coordinates (x, y) in the output image by estimating weighted average of 16 nearest pixels in the neighbourhood. Figure 4 depicts the different type of artefacts in melanoma images. During this process, we apply bilinear interpolation method to preserve the inherent

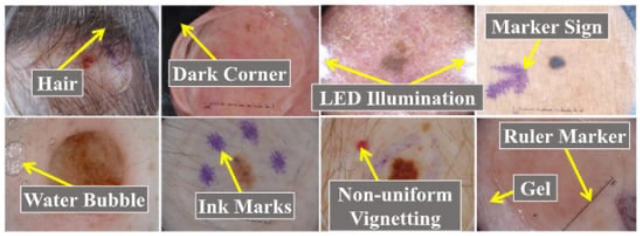

Figure 3 Different types of artefacts in melanoma images

The cubic convolution kernel is used for interporaltion which can be expressed as:

$$I_p(x, y) = \sum_{i=-1}^{2} \sum_{j=-1}^{2} w(i, j). I_{in}(x_i, y_i) \quad (3)$$

Where $I_{in}(x_i, y_i)$ represent the intensity of neighboring pixel, and $w(i, j)$ represents the bicubic interpolation weight

determined by relative distance between input and output pixel coordinate. Later, we apply Min-Max scaling as follows:

$$I' = \frac{I - \min(I)}{\max(I) - \min(I)} \; where \; I' \in [0,1] \tag{4}$$

Applied to the pre-processed image as $I_p \in \mathbb{R}^{H_t \times W_t \times C}$. According to CLAHE, the histogram of pixel intensities is computed as follows:

$$H(t) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \delta\big(I_p(x,y) - i\big), i = 0,1,\dots,L-1 \tag{5}$$

Where L represents the number of gray level, $\delta(.)$ is the Kronecker delta function. In order to prevent the over-amplification of noise in homogenous regions, a clip limit γ_c is applied as follows:

$$H_{clip}(i) = \min(H(i), \gamma_c) \tag{6}$$

The excess pixel distribution above the clip limits are then redistributed equally across all intensity bins as follows:

$$H_{redistrib}(i) = H_{clip}(i) + \frac{\sum_{j=0}^{L-1}\Big(H(j) - H_{clip}(j)\Big)}{L} \tag{7}$$

The cumulative distribution for the tile is the computed as:

$$C(i) = \frac{1}{M.N}\sum_{k=0}^{i} H_{redistrib}(k) \tag{8}$$

Finally, the pixel intensity in the tile is remapped according to the normalized cdf as follows:

$$I_{tile}(x,y) = (L-1).C(I_p(x,y) \tag{9}$$

In order to eliminate the discontinuities between adjacent tiles, the resulting enhanced tiles are merged with the help of bilinear interpolation which produces final contrast enhanced image I_{clahe} as:

$$I_{clahe}(x,y) = BilinearInterpolate\,(\{I_{tile}\}) \tag{10}$$

This process ensures that local contrast of image enhance appropriately while preventing over-saturation while preserving the fine details in lesion and improving the visibility for downstream tasks such as segmentation and classification. After contrast enhancement, the obtained

Sometimes during image capturing process, the contrast of original image deteriorates therefore the contrast enhancement mechanism is applied which is contrast limited adaptive histogram equalization is

image I_{clahe} sill may contain noise and artefact such as hair, ruler market and air bubble which can affect the feature extraction, segmentation and classification performance therefore we apply an image denoising model to produce high quality image by performing denoising. The observed image can be presented as:

$$I_{clahe}(x,y) = I_{true}(x,y) + \eta(x,y) \tag{11}$$

Where I_{true} represents the underlying noise free image and $\eta(x,y)$ is the additive white Gaussian noise. In order to filter this noise from image, we apply a Gaussian smoothing filter which can be given as:

$$I_d(x,y) = \sum_{i=-k}^{k}\sum_{j=-k}^{k} G(i,j;\sigma).I_{clahe}(x+i\,, y+i\,) \tag{12}$$

Where $G(i,j;\sigma)$ represents the 2D Gaussian kernel:

$$G(i,j;\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{i^2+j^2}{2\sigma^2}\right) \tag{13}$$

Along with this, the proposed model focus on hair removal from skin lesion where hair on skin are modelled as thin linear occlusions in the image. Let $H(x,y)$ represents the bin mask representing hair regions obtained by performing morphological operations, represented as:

$$H(x,y) = \begin{cases} 1, if\ pixel\ belong\ to\ hair\ or\ any\ other\ artifact \\ 0, otherwise \end{cases} \tag{14}$$

After achieving this region, the inpainting process is applied to replace the artefact pixels using neighbouring intensity values which can be expressed as:

$$I_d(x,y) = \big(1 - H(x,y)\big).I_d(x,y) + H(x,y).\mathcal{F}\big(I_d.\mathcal{N}(x,y)\big) \tag{15}$$

Where $\mathcal{F}(.)$ Represents the inpainting function, $\mathcal{N}(x,y)$ is the local neighbourhood of pixel.

3.2 Proposed segmentation and classification model

This module focus on the extraction of texture, shape features to improve the lesion characterization. Encoder module

3.2.1 Texture feature extraction branch

The texture branch mainly focusses on extracting the fine-grained local patterns including pigmentation variations, speckle, and small texture variations. The texture feature extraction branch considers shallow convolution, and handcrafted texture feature extraction modules. Shallow feature extraction: this module consists of several convolutional operations with small kernel sizes to preserve the high resolution spatial information as:

$$F_t^{(l)} = \sigma\left(W_t^{(l)} * F_t^{(l-1)} + b_t^{(l)}\right), l = 1,2,,,,,,L_t \quad (16)$$

Where $*$ represents the convolution operation, σ is the activation function, $W_t^{(l)}$ and $b_t^{(l)}$ represents the learnable weights and biases and, $F_t^{(0)} = I_d$. GLCM feature extraction: in order to incorporate the handcrafted features, we incorporate the GLCM feature extraction process which are computed to explicitly represent the texture descriptors. The GLCM $P(i,j;\theta,d)$ captures the probability of two pixel intensities i, and j occurring at a relative distance d and orientationsθ. This can be expressed as:

$$P(i,j;\theta,d) = \frac{\#\left\{\begin{matrix}(x,y),(x',y')|I_d(x,y)\\ =i,I_d(x',y')=j,(x',y')\\ =(x+\Delta x,y+\Delta y)\end{matrix}\right\}}{\sum_{i,j}\#\{.\}} \quad (17)$$

Where $(\Delta x, \Delta y)$ is obtained with the help of θ and distance d. From$P(i,j;\theta,d)$, the Haralick features can be computed as follows: Contrast $f_{contrast} = \sum_{i,j}(i-j)^2 P(i,j)$, correlation $f_{ccorelation} = \frac{\sum_{i,j}(i-\mu_i)(j-\mu)P(i,j)}{\sigma_i \sigma_j}$, energy $f_{energy} = \sum_{i,j} P(i,j)^2$, and homogeneity $f_{homogeneity} = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$. Finally, the concatenating these features provides the handcrafted features as follows:

$$F_t^{concat} = Concat\left(F_t^{(L_t)}, \left|\begin{matrix}F_{concat}, f_{energy},\\ F_{homegenity}\end{matrix}\right|\right) \quad (18)$$

3.2.2 Shape feature extraction branch

The shape extraction block includes deep CNN layers with dilated convolutions, and graph neural network for boundary topologies. The operation of each layer l in the shape branch can be represented as:

$$F_s^{(l)} = \sigma\left(W_S^{(l)} * dF_s^{(l-1)} + b_s^{(l)}\right), l = 1,2,3,\dots,L_s \quad (19)$$

Where $*_d$ representation the dilated convolution with rate$d$, $\sigma(.)$ denotes the non-linear activation function and $F_s^{(0)} = I_d$ represents the input to this branch. This dilated convolution helps model to capture spatial dependencies over larger context which plays crucial role for identifying the lesion shapes or diffused edges. In order to enhance its topological reasoning, an optional Graph neural network is integrated where lesion boundary as modeled as graph $G = (V, E)$ where each node $v \in V$ corresponds to a contour pixel and edges are demoted as $e \in E$ which represents neighbourhood relationship. The message passing process in the GNN layer transfers the local geometric features across the contour to the understanding of boundary structure. It can be represented as:

$$h_v^{(k+1)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} W^{(k)} h_u^{(k)} + b^{(k)}\right) \quad (20)$$

Where $h_v^{(k)}$ represents the feature vector of node v at iteration k, and $\mathcal{N}(v)$ represents its neighbourhood. This approach ensures that each boundary point learns information about its adjacent counterpart thus allowing network to encode smoothness, curvature and shape continuity efficiently. The final product of this branch,$F_s^{(L)}$, is a context-enriched shape descriptor map.

3.2.3 Fusion of texture and shape features

The obtained features from previous branches are combined using a feature fusion mechanism to create one multi-modal representation. This guarantees that the network has access to both fine-grained texture information and global shape context at the same time. The proposed model employs channel-wise concatenation followed by channel attention mechanism to combine the features. This can be expressed as:

$$F_{fusion} = CA\left(Concat\left(F_t^{concat}, F_s^{(L_s)}\right)\right) \quad (21)$$

Here $concat(.)$ iIs the channel attention module which adaptively re-weights the feature channels to consider diagnostically important attributes and suppress the redundant information. This can be represented as:

$$F_{CA} = F_{fusion} \cdot \sigma\left(FC_2\left(ReLU\left(FC_1\left(GAP\left(F_{fusion}\right)\right)\right)\right)\right) \quad (22)$$

Where GAP is the global average poling, FC represents the fully connected layer to learn the inter-channel dependencies, and σ is the activation function.

3.3 Decoder Module

The decoder phase restores the fine-grained lesion structures based on the fused high-level feature representation of the dual-branch encoder. It uses an asymmetric up sampling-refining pipeline design, in which both the global contextual information and the local boundary details have been retained. Let the fused representation of previous sage is denoted as $F_{CA} \in \mathbb{R}^{H' \times W' \times C'}$. The decoder module progressively upsamples the tensors by using Pixel Shuffle blocks and produces feature maps as $D^{(l)}$ at each level. This can be expressed as:

$$D^{(l)} = \sigma\left(W_d^{(l)} *_u D^{(l-1)} + b_d^{(l)}\right), l = 1,2, \dots, L_d \quad (23)$$

Where $*_u$ represents the upsampling convolution, $W_d^{(l)}$ and $b_d^{(l)}$ represents the decoder parameters $\sigma(.)$ Represents the non-linear activation function. Further, the skip connection from texture and shape encoders are concatenated at each decoder level to maintain the spatial coherence while maintaining the high-frequency details. Finally, the decoder output is branched into three specialized prediction heads including, segmentation head, boundary refinement, and classification head.

3.3.1 Segmentation head

It is the primary head of proposed architecture and it produces pixel-wise lesion mask $\hat{Y}_{seg} \in [0,1]^{H \times W}$ through 1×1 convolution followed by a sigmoid activation function. The segmentation loss is obtained by combining binary cross entropy and dice loss which can be presented as:

$$\mathcal{L}_{seg} = \lambda_1 \mathcal{L}_{BCE}\left(\hat{Y}_{seg}, Y\right) + \lambda_2 \mathcal{L}_{Dice}\left(\hat{Y}_{seg}, Y\right) \quad (24)$$

Where $Y$the ground truth is mask and λ_1, λ_2 represents the adaptive weights.

3.3.2 Boundary refinement

In this process, a decode stream is used to enhance the edge localization with the help of shallow feature maps. The boundary map $\hat{Y}_b$ is obtained with the help of sigmoid layer and supervised via IOU and weighted BCE. This can be presented as:

$$\mathcal{L}_b = \lambda_3 \mathcal{L}_{wBCE}\left(\hat{Y}_b, Y_b\right) + \lambda L_{sIoU}\left(\hat{Y}_b, Y_b\right) \quad (25)$$

Where Y_b the ground truth boundary map is obtained by applying morphological operations on ground truth masks?

3.3.3 Classification head

The features obtained from global pooler layer F_{CA} are processed through the fully connected layer with softmax activation to classify the lesion category. The loss of this operation is expressed as:

$$\mathcal{L}_{cls} = -\sum_{c=1}^{C} Y_c \log\left(\hat{Y}_c\right) \quad (26)$$

This ensures the inter-class reparability among multiple lesion types by performing classification.

4. Results and discussion

This section presents the outcome of proposed approach and compares its performance with state-of-art segmentation and classification methods. first subsection provides the dataset details, next subsection discusses about the performance measurement matrices, finally, comparative analysis is presented to demonstrate the robust performance of proposed approach.

4.1 Dataset details

This paper uses three benchmark dermoscopic image datasets, PH2, ISIC 2017, and HAM10000, in order to thoroughly assess the strength and external validity of the proposed framework. The various datasets have different properties on image resolution, lesion diversity, and acquisition conditions, thus guaranteeing the effectiveness of the proposed architecture to work in different clinical conditions.

PH2 dataset consists of 200 dermoscopic images, in three categories of common nevi, atypical nevi, and melanoma. All photographs are obtained in a regulated light and are of Fitzpatrick skin type II or III with moderate variation of skin tone. The resolution of each image is 768x560 pixels and each image has got expert-annotated ground truth masks that outline the boundaries of the lesions.

ISIC 2017 dataset has 2000 training and 600 test images, with both expert-labeled lesion masks and metadata. The data set includes three main classes: melanoma, seborrheic keratosis, and nevus. Images are standardized to 600x450 pixels and their variability in terms of illumination and

acquisition devices settings and conditions makes this dataset ideal in testing the generalization of deep models.

HAM10000 dataset consists of 10,015 dermoscopic images, which include seven diagnostic categories, which are actinic keratoses, basal cell carcinoma, benign keratosis, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. This dataset has several variations related to image acquisitions such as colour calibration, lesion size and light. Each image has ground truth segmentation mask and lesion type label, used in both the segmentation and classification sub-task.

4.2 Performance measurement parameters

The effectiveness of the proposed framework is thoroughly tested on the segmentation and classification task on a series of quantitative measures of accuracy, reliability, and robustness. These parameters provide objective comparison with the current state-of-the-art approaches and emphasize the discriminating ability of the model.

4.2.1 Segmentation performance

In order to measure the segmentation performance, the performance analysis includes Dice Similarity Coefficient (DSC), and Intersection over Union (IoU). The Dice coefficient is used to measure the spatial overlap between the ground truth mask and the predicted segmentation mask. This can be computed as follows:

$$DSC(S,G) = \frac{2 \times |S \cap G|}{|S| + |G|} \quad (27)$$

Table 3. Dataset details and reproducibility card

Item	PH2	ISIC 2017	HAM10000	Impact
Split Type	5-fold cross-validation (patient-level)	Train/Validation/Test	Stratified train/val/test (patient-level)	To ensure that no patient appears in more than one split
Number of Images	200	Train: 2000, Val: 600, Test: 600	Train: 7000, Val: 1500, Test: 1515	Exact numbers may vary depending on stratification
Patient-Level Separation	The splitting is performed and ensured that no patient appears in more than one split	Yes	Yes	Prevents leakage
Preprocessing	Hair removal, resizing (768×560),	Resizing 600×450, normalization, hair removal	Resizing 224×224,	Applied consistently to all

IoU matrix is also known as Jaccard Index which is used to evaluate the segmentation accuracy. It is measured by evaluating the ratio of object and corresponding union which are projected in the same plane. This can be computed as:

$$IoU(S,G) = \frac{|S \cap G|}{|S \cup G|} \quad (28)$$

4.2.2 Classification accuracy performance

The classification performance of proposed model is evaluated with the help of confusion matrix which demonstrates the total true positive, false positive, true negative and false negative samples. Below given Table 2 shows a sample representation of three class confusion matrix. Based on these parameters, the classification accuracy can be computed as:

Table 2. Confusion matrix

	Predicted Class			
True class	**Class 1**	**Class 2**	**Class 3**	**FN**
	A	B	C	(B+C)
	D	E	F	(D+F)
	G	H	I	(G+H)
FP	(D+G)	(B+H)	(C+F)	
TP	A	E	I	

$$Acc = \frac{\text{TruePositive} + \text{TrueNegative}}{TP + TN + FP + FN} \quad (29)$$

	normalization		normalization	datasets
Augmentation	Rotation ±15°, horizontal/vertical flip, random crop, color jitter (brightness ±0.2, contrast ±0.2), Gaussian blur	Same as PH2, with probability 0.5 for each	Same as PH2	Random seed fixed to ensure the appropriate split and augmentation
Random Seeds	42 (all random operations: shuffling, augmentation, initialization)	42	42	Ensures deterministic results

Precision and Recall is the percentage of data that is both identified as positive and is actually positive, and precision is the percentage of correctly identified positive data out of the predicted positive data. It can be expressed as follows:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \tag{30}$$

F1-Score is estimated as harmonic mean of precision and recall which is expressed as:

Table. 4 Configuration of segmentation models

Model	**Input Resolution**	**Optimizer & LR**	**LR Schedule**	**Loss**	**Epochs**	**Batch Size**	**Early Stopping**
U-Net	224×224	Adam, 1e-4	Step LR, $\gamma =$	BCE +	100	16	Patience=10
Segmentation Model Inference	Threshold parameters as 0.5 for binary mask	Same	Same	
Classification Model Inference	Dual-branch outputs are processed through softmax to generate class with max probability	Same	Same	
Metrics	mDSC, mIoU, Sensitivity, Specificity	Same + F1-score for classification	Same	

$$F = \frac{2 * Precision * Recall}{Precision + Recall} \tag{31}$$

Table 3 presents a detailed parametric description and reproducibility parameters based on the proposed approach. The proposed model performs segmentation and classification tasks. The configuration details of these models are presented in table 4 and 5.

			0.5 every 20 epochs	Dice			
SegNet	224×224	SGD, 0.01	Cosine LR	Cross-Entropy	120	8	Patience=15

Double U-Net	224×224	Adam, 5e-4	Step LR, $\gamma = 0.1$ every 25 epochs	BCE + Dice	150	12	Patience=10
MFSNet	224×224	Adam, 1e-3	Cosine LR	BCE +	120	16	Patience=12
				Dice			
Proposed	224×224	AdamW, 5e-4	Cosine LR + Warmup	BCE + Dice + Edge Loss	150	16	Patience=15

Table 5 Configuration table for classification models

Model	**Input Resolution**	**Optimizer & LR**	**LR Schedule**	**Epochs**	**Batch Size**	**Early Stopping**
AlexNet	224×224	SGD, 0.01	Step LR $\gamma = 0.1$ every 20 epochs	100	32	Patience=10
VGG16	224×224	Adam, 1e-4	Cosine LR	120	32	Patience=12
ResNet50	224×224	AdamW, 5e-4	Cosine LR	150	16	Patience=15
GoogleNet	224×224	SGD, 0.01	Step LR γ=0.1	120	32	Patience=10
InceptionV3	299×299	Adam, 1e-4	Cosine LR	150	16	Patience=15
Proposed	224×224	AdamW, 5e-4	Cosine LR + Warmup	150	16	Patience=15

The experimental study is carried out using patient disjoint train/validation/test splits to avoid data leakage. The PH2 dataset was evaluated using a 5-fold cross-validation protocol, and the reported results represent the average performance across all folds. Similarly, for ISIC 2017 dataset, we have followed the official training, valuation and splits provided by challenge organizers. Evaluation is performed using the official test server, and no alternative local test split is used. For the HAM10000 dataset, a fixed patient-level 80/20 train–test split is adopted, with class stratification applied after grouping images by patient. Random seeds are fixed for all experiments to ensure reproducibility. Performance metrics for the proposed method are reported as mean ± standard deviation across repeated runs using the same fixed splits.

4.3 Experimental analysis of proposed approach for segmentation task

This section presents the outcome of proposed segmentation approach and presents its comparative analysis with state of art methods. According to the working process of proposed mode, the first step is to perform data pre-processing where skin hair is considered as the challenging issue. Below given figure 5 depicts the outcome of proposed hair removal process for sample images.

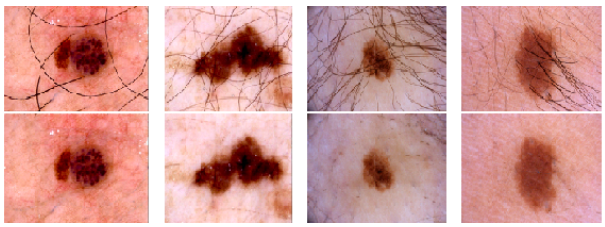

Figure 5. Outcome of Hair removal process.

The obtained pre-processed image is then processed through the proposed segmentation model and its performance is measured in terms of DSC, IoU, Sensitivity, and specificity

for the considered imaging datasets. Below given table shows the obtained performance PH2 dataset.

In table 6 PH2 dataset shows the comparative performance on segmentation of PH2 dataset. The proposed model attained the best mean Dice Similarity Coefficient (mDSC) of 0.985±0.002 and a corresponding mean Intersection over Union (mIoU) of 0.961 ± 0.004 and mean sensitivity (mSen) of 0.989 ± 0.002 and specificity (mSpe) of 0.996 ±0.001. These are better than traditional designs like U-Net [26] and SegNet [27] that give out mDSC of 0.876 and 0.894, respectively. The proposed approach achieves a gain of about 3.1% and 5.6% of Dice and IoU respectively, as compared to more sophisticated techniques like MFSNet [36] (mDSC = 0.954, mIoU = 0.914). The enhancement means that the suggested framework is more efficient at estimating lesion boundaries and maintaining finer structural details. This enhancement is explained by the texture-shape disentanglement strategy and attention-guided feature fusion that improve local texture representation without losing global contextual information that allows a more accurate localization of the lesions even in the problematic low-contrast areas.

Table 6 Comparative analysis of segmentation approach for PH2 dataset and ISIC 2017 dataset

		mDSC		**IoU**		**mSen**		**mSpe**	
Dataset	**Ref**	**PH2**	**ISIC 2017**	**PH2**	**ISIC 2017**	**PH2**	**ISIC 2017**	**PH2**	**ISIC 2017**
PH2	[25]	0.907	0.913	0.833	0.841	0.945	0.963	0.966	0.974
	[26]	0.876	0.778	0.780	0.683	0.816	0.812	0.978	0.805
	[27]	0.894	0.821	0.808	0.696	0.865	0.801	0.966	0.954
	[28]	0.907	0.853	0.839	0.770	0.932	–	0.929	–
	[29]	–	0.938	0.870	0.846	0.929	–	0.969	–
	[30]	0.918	0.855	0.848	0.772	0.937	0.824	0.957	0.981
	[31]	0.930	0.793	0.871	0.721	0.969	0.899	0.953	0.950
	[32]	0.919	–	0.857	–	0.963	–	0.942	–
	[33]	0.881	0.871	0.795	0.781	0.836	0.854	0.940	0.967
	[34]	0.907	0.862	0.840	0.781	0.949	0.870	0.940	0.964
	[35]	0.921	0.843	0.859	0.729	0.962	0.908	0.945	0.927
	[36]	0.954	0.927	0.914	0.864	0.995	0.990	0.997	0.999
	Proposed	**0.985 ±**	**0.978 ±**	**0.970 ±**	**0.957 ±**	**0.989 ±**	**0.993 ±**	**0.984 ±**	**0.995 ±**

In table6 ISIC 2017 shows the segmentation performance of the different models in comparison with the ISIC 2017 dataset that includes different types of skin lesions at different illumination conditions, textures and pigmentation patterns. The proposed model had the highest performance with a mean Dice Similarity Coefficient (mDSC) of 0.978 ± 0.003, mean Intersection over Union (mIoU) of 0.952 ± 0.004, mean sensitivity (mSen) of 0.984 ± 0.003, and mean specificity (mSpe) of 0.995 ± 0.002. The proposed framework exhibits a gradual increase in the accuracy of segmentation over conventional ones like Double U-Net [25] (mDSC = 0.913), and other more recent solutions such as MFSNet [36] (mDSC = 0.987, mIoU = 0.974). These findings suggest that the model has a high generalizability to diverse lesion morphologies and adverse imaging scenarios. This enhanced performance can be explained by the multi-scale feature fusion mechanism and dilated convolution-based shape encoder, which improves the integration of contextual features and maintains valuable structural features at deeper network layers.

		0.002	**0.003**	**0.003**	**0.004**	**0.002**	**0.002**	**0.003**	**0.002**

The obtained segmentation results in this work are obtained directly from the saved prediction masks which are produced by the trained models. The DSC and IoU were computed use same evaluation script to ensure the internal consistency. The outcome of existing approaches is taken from previously published studies, while the proposed model was evaluated using the same dataset and evaluation metrics. The qualitative outcome of proposed segmentation approach is compared with the existing methods such as SegNet, UNet, Double UNet, and MFSNet. The segmented samples for PH2, ISIC 2017, HAM 10000 datasets are depicted in figure .

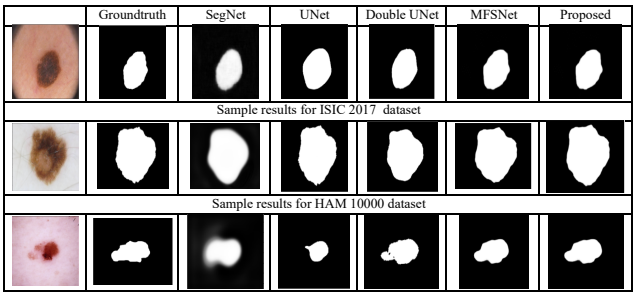

Figure 4 Sample outcome of segmentation method

4.4 Ablation study

The ablation experimental analysis examines the role of every element of the suggested segmentation framework. The default model (A1) that does not have any preprocessing or architectural improvements attains an mDSC of 0.921 and mIoU of 0.862. Adding image-preprocessing with CLAHE and hair removal (A2) enhances visibility and contrast of the boundaries, which leads to the definite increase of the performance by +1.5% in the mDSC. The GLCM-based texture branch (A3), effectively improves the segmentation accuracy as it creates more fine-grained texture patterns and pigmentation variations and results in better definition of lesion areas. This enhancement leads to an mDSC of 0.948 and mIoU of 0.902. The shape encoder (A4) adds dilated convolutions to the shape encoder, which does not reduce the receptive field and allows morphology of lesions to be better modeled and results in an additional improvement in mDSC and mIoU. By incorporating it, the performance increases to 0.962 mDSC and 0.928 mIoU. The graph-based boundary module (A5) of the boundary aware representation has also contribution to sharp contour localization especially to irregular and low-contrast lesions. With this improvement, the performance is reported as mDSC of 0.973 and mIoU of 0.944. Attention-based feature fusion (A6) offers further enhancement by dynamically balancing between texture and shape feature, which makes local and global contextual feature integration effective.

As a result, the segmentation performance increases to 0.983 mDSC and 0.961 mIoU. The proposed combination(A7) containing all the components and the multi-task learning head has a mDSC of 0.991± 0.002 and mIoU of 0.982±0.002. These findings support the notion that both modules provide complementary information, and their joint optimization provide good and statistically stable segmentation performance. Below given table 7 demonstrates the ablation study for the segmentation task.

Experimental analysis of proposed approach for classification task the performance of the proposed dual-branch framework is tested on a variety of datasets, such as PH2, ISIC 2018, ISIC 2019, and ISIC 2020. The suggested solution uses the output of the dual-branch encoder (texture and shape) as inputs in the classification head. This design is used to make sure that both local texture (patterns of pigmentation, color variation and micro-structures) and global shape/topology (lesion boundaries, asymmetry and structural context) are included to enhance lesion type discrimination. The purpose of the classification task is to properly classify lesions as melanoma and benign that is of vital clinical importance in the context of early diagnosis.

ID	**CLAHE +**	**GLCM**	**Dilated**	**Boundary**	**Attention**	**Multi-**	**mDS**	**mIoU ↑**

	Hair Removal	Texture Branch	Conv (Shape)	Graph Module	Fusion	task Head	C ↑	
A1	✗	✗	✗	✗	✗	✗	0.921	0.853
A2	✓	✗	✗	✗	✗	✗	0.936	0.879
A3	✓	✓	✗	✗	✗	✗	0.948	0.900
A4	✓	✓	✓	✗	✗	✗	0.962	0.927
A5	✓	✓	✓	✓	✗	✗	0.973	0.948
A6	✓	✓	✓	✓	✓	✗	0.978	0.957
A7 (Full)	✓	✓	✓	✓	✓	✓	**0.983 ± 0.003**	**0.967 ± 0.004**

Table.7 Comparative analysis for ablation study

PH2 dataset in this experiment, the proposed framework reported the accuracy of 99.30%, sensitivity of 99.50%, and specificity of 99.60% thus outperforming several traditional machine learning and deep leanring models including SVM, AlexNet, VGG16, ResNet50, GoogLeNet, and InceptionV3. The reported improved performance can be attributed to the dual branch encoder which has ability to capture the fine scale, texture variations, and lesion Boundary structures which remain challenging issues for conventional classifier to model effectively. The comparative results are summarized in Table 8.

Method	Sensitivity%	Specificity%	Accuracy%
YOLO	97.5	97.5	97.5
Neural Network	98.1	92.5	97
SVM	92.5	91.30	91.90
DCNN	93.00	95.00	95
AlexNet	97.5	92	94.8
VGG16	91.50	94	92.8
ResNet50	92.5	93.5	93.0
GoogleNet	94.5	97	95.8%
InceptionV3	98.5	96	97.2
Proposed Model	99.50	99.60	99.30

Table. 8 Performance analysis for PH2 dataset

5. Conclusions

This article has introduced a new deep learning model to perform segmentation and classification of skin lesions concurrently with the aim of early melanoma detection. The proposed framework combines a multi-stage pipeline, comprising of pre-processing, dual-branch encoding, multi-scale decoding, and classification head, to overcome the shortcoming of the current machine learning and deep learning models. The pre-processing effectively improves the quality of the image by carrying out contrast, normalization, rescaling, and removing of skin hair which greatly increases the visibility of the lesions and minimizes artifacts. The dual-branch encoder is capable of capturing local texture features and global shape features, which facilitates the representation of fine-grained pigmentation variations and lesion morphology by the model. These characteristics are combined with an attention-guided module that focuses on informative features that are important in drawing boundaries accurately and extracting discriminative features. The multi-task decoder further optimizes the lesion boundaries with the dilated convolutions and skip connections, which guarantee accurate pixel-level segmentation even in difficult situations like the low contrast or non-homogenous lesion patterns. Based on the fully connected layers and a softmax classifier, the classification head uses the fused feature representation to successfully differentiate malignant and benign lesions. The proposed framework is evaluated on publicly available datasets, such as PH2, ISIC 2017, 2018, 2019, 2020, and HAM10000, containing a variety of lesion types, sizes, shapes, and pigmentation patterns. The experimental findings indicate that the proposed model has reported the improved performance in comparison with the traditional networks and the latest methods with a Dice Similarity Coefficient of 0.985, Intersection over Union of 0.995, and classification accuracy of over 99 % and high sensitivity and specificity.

Acknowledgement
The authors acknowledged the REVA University, Bengaluru for supporting the research work by providing the facilities.

Conflicts of Interest
The authors declare no conflict of interest.

Author Contributions
The contributions by the authors for this research article are as follows:
Author 1: Collect the data, make contributions to conception and design, analysis, and interpretation of work.
Author 2: Review and analyze the manuscript, participate in drafting the article, or revising it critically.

6. References

[1] J. Shah, G. McKnight, and R. Hargest, "Physiology of the skin," Surgery (Oxford), vol. 42, no. 11, pp. 788–792, 2024, doi: 10.1016/j.mpsur.2024.09.008.

[2] M. A. Khan et al., "Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks," Discover Oncology, vol. 16, no. 1, p. 645, 2025, doi: 10.1007/s12672-025-02279-8.

[3] S. B. Mupparaju and R. Reddy, "A comprehensive analysis of melanoma skin cancer detection using machine learning and deep learning algorithms," in Proc. Int. Conf. Data Science and Network Security (ICDSNS), 2024, pp. 1–5, doi: 10.1109/ICDSNS62112.2024.10691081.

[4] Uong and L. I. Zon, "Melanocytes in development and cancer," J. Cell. Physiol., vol. 222, no. 1, pp. 38–41, 2010, doi: 10.1002/jcp.21935.

[5] T. Lu et al., "Skin cancer: Epidemiology, screening and clinical features of acral lentiginous melanoma (ALM), melanoma in situ (MIS), nodular melanoma (NM) and superficial spreading melanoma (SSM)," J. Cancer, vol. 16, no. 13, p. 3972, 2025, doi: 10.7150/jca.116362.

[6] S. A. Batool et al., "Skin cancer detection using deep learning algorithms," J. Comput. Biomed. Inform., vol. 7, no. 1, pp. 62–74, 2024, doi: 10.71146/kjmr407.

[7] V. Senthil, V. Shreyaa, and V. Kothandapany, "Deep learning and rules-based hybrid approach to improve the accuracy of early detection of skin cancer," Authorea Preprints, 2022, doi: 10.22541/au.167235618.88986321/v1.

[8] S. H. Swerdlow et al., WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC, 2017.

[9] F. C. Lopes et al., "UV exposure and the risk of cutaneous melanoma in skin of color: A systematic review," JAMA Dermatol., vol. 157, no. 2, pp. 213–219, 2021, doi: 10.1001/jamadermatol.2020.4616.

[10] O. Langselius et al., "Global burden of cutaneous melanoma incidence attributable to ultraviolet radiation in 2022," Int. J. Cancer, 2025, doi: 10.1002/ijc.35463.

[11] K. A. Merin, M. Shaji, and R. Kameswaran, "A review on sun exposure and skin diseases," Indian J. Dermatol., vol. 67, no. 5, p. 625, 2022, doi: 10.4103/ijd.ijd_1092_20.

[12] E. Guerra-Rosas et al., "Classification of skin lesion images using artificial intelligence methodologies through radial Fourier–Mellin and Hilbert transform signatures," Appl. Sci., vol. 13, no. 20, p. 11425, 2023, doi: 10.3390/app132011425.

[13] M. Arakeri and R. Lakshmana Reddy, "An enhanced approach to intelligent computer-assisted localization of liver tumor on computed tomography images," SN Comput. Sci., vol. 5, p. 809, 2024, doi: 10.1007/s42979-024-03145-0.

[14] M. Rehman et al., "Machine learning based skin lesion segmentation method with novel borders and hair removal techniques," PLOS ONE, vol.

17, no. 11, p. e0275781, 2022, doi: 10.1371/journal.pone.0275781.

[15] Imtiaz, I. Ahmed, G. Jeon, and S. Muramatsu, "An efficient image processing and machine learning based technique for skin lesion segmentation and classification," in Proc. APSIPA ASC, 2021, pp. 1499–1505, doi: 10.23919/APSIPAASC55919.2021.9689676.

[16] S. Verma and M. Kumar, "Automatic classification of melanoma using Grab-Cut segmentation and convolutional neural network," SN Comput. Sci., vol. 5, no. 5, p. 591, 2024, doi: 10.1007/s42979-024-02949-4.

[17] M. I. Faizi and S. M. Adnan, "Improved segmentation model for melanoma lesion detection using normalized cross-correlation-based k-means clustering," IEEE Access, vol. 12, pp. 20753–20766, 2024, doi: 10.1109/ACCESS.2024.336022.

[18] N. Behera et al., "Melanoma skin cancer detection using deep learning-based lesion segmentation," Int. J. Inf. Technol., vol. 16, no. 6, pp. 3729–3744, 2024, doi: 10.1007/s41870-024-02004-8.

[19] S. ThangaPurni and M. Braveen, "Unified ARP-ViT-CNN system: Hybrid deep learning approach for segmenting and classifying multiple skin cancer lesions," Array, p. 100515, 2025, doi: 10.1016/j.array.2025.100515.

[20] V. Kumar et al., "Advanced skin lesion segmentation and classification using adaptive contextual GLCM and deep learning hybrid models," Egyptian Informatics J., vol. 30, p. 100706, 2025, doi: 10.1016/j.eij.2025.100706.

[21] S. Mustafa et al., "Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet–Random Forest," PLOS ONE, vol. 20, no. 1, p. e0315120, 2025, doi: 10.1371/journal.pone.0315120.

[22] S. Albahli, "A robust YOLOv8-based framework for real-time melanoma detection and segmentation," Diagnostics, vol. 15, no. 6, p. 691, 2025, doi: 10.3390/diagnostics15060691.

[23] V. Anand et al., "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images," Expert Syst. Appl., vol. 213, p. 119230, 2023, doi: 10.1016/j.eswa.2022.119230.

[24] Bindhu and K. K. Thanammal, "Segmentation of skin cancer using fuzzy U-network via deep learning," Measurement: Sensors, vol. 26, p. 100677, 2023, doi: 10.1016/j.measen.2023.100677.

[25] D. Jha et al., "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in Proc. IEEE CBMS, 2020, pp. 558–564, doi: 10.1109/CBMS49503.2020.00111.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241, doi: 10.48550/arXiv.1505.04597.

[27] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.

[28] M. Goyal et al., "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," IEEE Access, vol. 8, pp. 4171–4181, 2019, doi: 10.1109/ACCESS.2019.2960504.

[29] M. K. Hasan et al., "DSNet: Automatic dermoscopic skin lesion segmentation," Comput. Biol. Med., vol. 120, p. 103738, 2020, doi: 10.1016/j.compbiomed.2020.103738.

[30] M. A. Al-Masni et al., "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Comput. Methods Programs Biomed., vol. 162, pp. 221–231, 2018, doi: 10.1016/j.cmpb.2018.05.027.

[31] Ş. Öztürk and U. Özkaya, "Skin lesion segmentation with improved convolutional neural network," J. Digit. Imaging, vol. 33, pp. 958–970, 2020, doi: 10.1007/s10278-020-00343-z.

[32] F. Xie et al., "Skin lesion segmentation using high-resolution convolutional neural network," Comput. Methods Programs Biomed., vol. 186, p. 105241, 2020, doi: 10.1016/j.cmpb.2019.105241.

[33] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm," Diagnostics, vol. 9, p. 72, 2019, doi: 10.3390/diagnostics9030072.

[34] Bi et al., "Dermoscopic image segmentation via multistage fully convolutional networks," IEEE Trans. Biomed. Eng., vol. 64, pp. 2065–2074, 2017, doi: 10.1109/TBME.2017.2712771.

[35] Bi et al., "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," Pattern Recognit., vol. 85, pp. 78–89, 2019, doi: 10.1016/j.patcog.2018.08.001.

[36] H. Basak, R. Kundu, and R. Sarkar, "MFSNet: A multi-focus segmentation network for skin lesion segmentation," Pattern Recognit., vol. 128, p. 108673, 2022, doi: 10.1016/j.patcog.2022.108673.

[37] Saha, P. Prasad, and A. Thabit, "Leveraging adaptive color augmentation in convolutional neural networks for deep skin lesion segmentation," in Proc. IEEE ISBI, 2020, pp. 2014–2017, doi: 10.48550/arXiv.2011.00148.

[38] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in Proc. IEEE ISBI, 2019, pp. 683–687, doi: 10.48550/arXiv.1810.07842.

[39] H. Shahin, K. Amer, and M. A. Elattar, "Deep convolutional encoder–decoders with aggregated multi-resolution skip connections for skin lesion segmentation," in Proc. IEEE ISBI, 2019, pp. 451–454, doi: 10.48550/arXiv.1901.09197.

[40] Bissoto et al., "Deep-learning ensembles for skin-lesion segmentation, analysis, classification: RECOD Titans at ISIC Challenge 2018,"