

# Adaptive Swarm-Driven Skin Lesion Segmentation and Classification Using Boundary-Aware U-Net and Hybrid Vision Transformer

Thilagavathi Rajamanickam<sup>1</sup>, Anuradha Konidena<sup>2</sup>, N. Mahesh<sup>3</sup>, M. Balakrishnan<sup>4</sup>,  
N. K. Priyadharsini<sup>5</sup>, Suganthi Rajamanickam<sup>6</sup>

<sup>1</sup> Department of Mathematics, Rajalakshmi Engineering College, Chennai, India.

<sup>2</sup> Department of CSE, Koneru Lakshmiah Education Foundation (KLEF), (Deemed to be University), Vaddeswaram, Guntur, Andra Pradesh, India.

<sup>3</sup> Department of Electronics and Instrumentation Engineering, Kongu Engineering College, Perundurai Erode, India

<sup>4</sup> Department of Artificial Intelligence and Data Science, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, India

<sup>5</sup> Department of Computer Science Engineering, P. A. College of Engineering and Technology, Pollachi, Coimbatore, India.

<sup>6</sup> Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, India.

**Corresponding author:** Dr. Thilagavathi Rajamanickam (e-mail: [thilagamsc@gmail.com](mailto:thilagamsc@gmail.com)), Author(s) Email: Anuradha Konidena (e-mail: [akonidena75@gmail.com](mailto:akonidena75@gmail.com)), Dr.N.Mahesh (e-mail: [maheshmeae@gmail.com](mailto:maheshmeae@gmail.com)), Dr.M.Balakrishnan (e-mail: [balakrishnanme@gmail.com](mailto:balakrishnanme@gmail.com)), N.K.Priyadharsini (e-mail: [priyadharsini.pacet@gmail.com](mailto:priyadharsini.pacet@gmail.com)), Suganthi Rajamanickam (e-mail: [sugimanicks@gmail.com](mailto:sugimanicks@gmail.com))

**Abstract** Skin cancer is one of the most prevalent dermatological diseases worldwide, and timely and accurate diagnosis is essential for effective treatment and improved patient outcomes. However, automated skin lesion analysis remains challenging due to irregular lesion boundaries, low contrast between lesions and surrounding skin, imaging artifacts, variations in lesion appearance, and significant intra-class variability. To address these challenges, this research proposes an Adaptive Swarm-Driven Skin Lesion Segmentation and Classification Approach that integrates an Adaptive Boundary-Aware U-Net (ABAU-Net), a Butterfly Optimization Algorithm–Elephant Herding Optimization (BOA-EHO) strategy, and a Hybrid Vision Transformer with Dynamic Feature Fusion (HViT-DFF). The proposed ABAU-Net is designed to enhance segmentation accuracy by emphasizing boundary information and preserving fine-grained lesion structures, particularly in cases involving poorly defined or irregular borders. The BOA-EHO optimization strategy combines global exploration and local exploitation mechanisms to improve network parameter optimization, accelerate convergence, and reduce the possibility of premature convergence. Following segmentation, the HViT-DFF model integrates CNN-based local spatial features with transformer-based global contextual representations through dynamic feature fusion, enabling robust discrimination among multiple skin lesion classes. Extensive experiments conducted on the ISIC skin lesion dataset demonstrate the effectiveness of the proposed framework, achieving a Dice coefficient of 98.2%, Jaccard index of 97.5%, classification accuracy of 98.8%, and AUC of 0.992. These results outperform recent CNN-, Transformer-, and hybrid-based approaches. Overall, the proposed framework demonstrates strong potential for accurate, reliable, and automated skin lesion segmentation and multi-class classification, supporting computer-aided dermatological diagnosis and potentially facilitating earlier clinical intervention.

**Keywords** Skin Lesion Segmentation, Boundary-Aware U-Net, Hybrid Vision Transformer, Swarm Optimization, Dermoscopic Image Classification.

## 1. Introduction

Skin cancer is the most common cancer in the world and melanoma is the most dangerous because of its

ability to metastasize. A timely diagnosis greatly improves patient survival. Dermoscopy is considered as the most valuable tool for imaging skin lesions as it

gives information about the morphology, pigmentation and borders of the lesions. Correct interpretation is, however, difficult because of different manifestations of the lesions, differences in illumination, the effects of hair interfacing, differences in image quality and inter-observer variability between dermatologists [1-3]. The automated analysis of skin lesions has made great strides in the past few years thanks to the progress in the field of deep learning. The encoder-decoder architecture based on U-Net has been widely used for lesion segmentation tasks and has shown excellent performance, while the Convolutional Neural Network (CNN) and Vision Transformer (ViT) have been used to extract global context information and local texture information, respectively, and have demonstrated great potential in skin lesion classification [4-12]. However, current segmentation approaches are hard to maintain non-regular boundaries of lesions, and CNNs have not been able to model global context and transformer-based models have large-scale requirements and missed the fine texture information [13-15].

This work proposes an Adaptive Swarm-Driven Skin Lesion Analysis Framework (ASSLA), which integrates three major components: (i) Adaptive Boundary-Aware U-Net (ABAU-Net) to obtain the accurate segmentation results while preserving the skin lesion boundaries, (ii) a Hybrid Butterfly Optimization Algorithm–Elephant Herding Optimization (BOA-EHO) strategy to enhance the convergence and optimize the training of the segmentation model, and (iii) a Hybrid Vision Transformer with Dynamic Feature Fusion (HViT-DF) to effectively fuse the features of the convolutional network and the Transformer network for robust multi-class skin lesion classification.

### A. Clinical Problem

One of the most common types of cancer worldwide is skin cancer, and melanoma is the most lethal type of skin cancer owing to the highly metastatic nature of the tumor. Accurate and early diagnosis plays a vital role in ensuring successful treatment. Dermoscopy is a well-known imaging modality that is widely used for diagnosing skin lesions; nevertheless, manual diagnosis can be difficult due to a number of factors.

### B. State-of-the-Art Segmentation and Classification Methods

The recent development of deep learning has greatly enhanced automatic skin lesion segmentation. The U-Net architecture and its variations have dominated the field of lesion segmentation due to their efficiency in capturing multi-scale contextual information. Likewise, the Convolutional Neural Network (CNN) model has shown promising results in skin lesion classification through the ability to learn discriminative texture and shape features. Vision Transformers (ViT) and CNN-Transformers have been recently developed to better

learn long-range contextual dependency for lesion classification.

### C. Research gap:

Despite the fact that recent frameworks for skin lesions' analysis, which use CNNs, Vision Transformers, and hybrid architectures of CNN and Transformer models, show high segmentation and classification accuracy, there are still some critical issues to be resolved. Current segmentation models face difficulties in preserving boundaries of irregular shape and fine details of lesion structure when dealing with low-contrast dermoscopic images that suffer from noise, illumination and hair artifacts. Moreover, most classification frameworks are unable to use information about local textures and global dependencies, thus failing to distinguish visually similar types of lesions. All current solutions are also based on the traditional gradient-based optimization algorithms, which can suffer from slow convergence, local optima, and unstable training process. Current feature fusion approaches are mostly static and do not utilize complementary information contained in CNN and Transformer representations. Thus, it is necessary to design an integrated framework for improved boundary-aware segmentation, optimization, feature fusion, and classification of skin lesions.

## II. State-of-the-Art Techniques

Detection of skin lesions through automatic analysis has gained significant attention in recent years owing to the rise in cases of skin cancer such as melanoma [16]. The early detection methods involved the use of manually engineered features together with conventional machine learning algorithms in the analysis and segmentation of lesions [17]. The manually engineered features comprised texture descriptors, color histograms, edges, shapes, and morphology among others [18]. Despite the ease of implementation and simplicity compared to modern deep learning models, traditional models were highly sensitive to feature extraction and images themselves. Moreover, the traditional methods were limited in generalizing different cases involving different appearances of lesions, illumination levels, hair, and boundary shapes [19].

In order to tackle these problems, deep learning-based convolutional neural network (CNN)-based approaches were widely used for the automated segmentation and classification of skin lesions [20]. Frameworks that included deep neural feature fusion techniques and feature selection methods increased the lesion discrimination accuracy through the use of multi-scale representations and deep semantics [21]. Moreover, two-stage deep CNN-based frameworks utilizing weakly supervised segmentation methods led to higher efficiency for lesion localization and

classification without reliance on manual annotations [22]. Light weight deep CNN-based models were also proposed to decrease computational burden and increase practicality for real-time applications in clinic [23]. However, the traditional CNN-based approaches are still restricted in precisely capturing lesion borders and modeling long-range contextual relations in dermoscopic images.

Recently, boundary aware segmentation frameworks were suggested to improve lesion edge detection and avoid data loss at lesion boundaries. For instance, Amin et al. [12] designed a boundary aware segmentation architecture with the help of hybrid CNN-transformer classification technique. In the same vein, robust segmentation techniques based on deep networks were developed for enhancing robustness of lesion segmentation under noisy imaging conditions [24]. Confidence-guided ConvLSTM segmentation framework also helped improve lesion boundary consistencies through curriculum-based training and temporal feature learning [25]. Further, LEDNet and Swin-UMamba-based hybrid segmentation techniques were also utilized for boosting feature propagation and contextual learning for lesion segmentation [26]. Despite obtaining good segmentation results, many state-of-the-art frameworks continue facing issues related to poor boundary preservation, low contrast lesion robustness, and training time instabilities [27].

Hybrid CNN-transformers have shown an improvement in the segmentation and classification capabilities of the system by integrating convolutional neural networks with self-attention techniques [28], [29]. The integration of vision transformers into ResNet architectures and long short-term memory networks resulted in better sequential modeling and lesion classification capabilities [30]. In addition, multimodal learning frameworks were proposed to include diverse dermatoscopic data for early detection of skin cancer [31]. Hybrid deep learning models based on explainability performed clinical interpretation along with segmentation and multi-class classification tasks [9]. Foundation model approaches that leveraged SAM adapters and vision transformers further enhanced generalizability through large-scale pre-training and fine-tuning [32]. Lastly, attention-driven deep learning models have been employed for large-scale multi-class skin lesion classification tasks involving many classes [33].

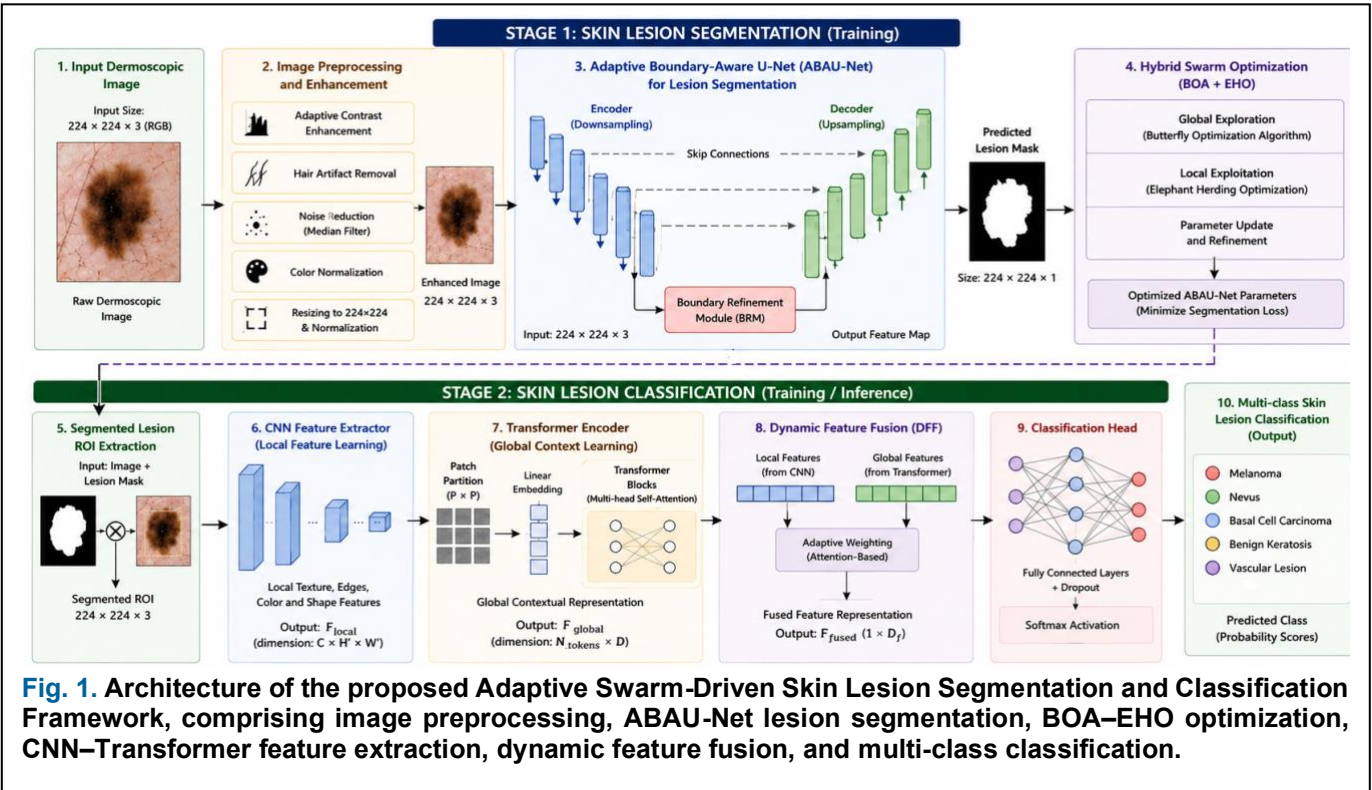
While transformer-based methods enhance contextual learning, they tend to necessitate massive amounts of training data and heavy computation capabilities [34]. Hybrid models are mostly characterized by static feature combination techniques incapable of dynamically responding to lesion variation and similarities. Therefore, developing an efficient and robust framework for combining both local and global feature representations constitutes one of the most

pressing research areas [35]. Optimization-driven and adaptive hybrid systems have been utilized in recent years to optimize convergence and boost diagnostic efficacy. Adaptive hybrid AI frameworks with robustness and explainability capabilities proved to be effective in detecting melanomas [36]. Confidence-weighted semi-supervised learning models based on hybrid CNN-transformer systems significantly boosted the robustness of lesion segmentation when dealing with scarce annotation conditions [37]. Robust lesion prediction models leveraging hybrid deep learning frameworks employed adaptive learning methodologies to achieve better generalizability [38]. Nevertheless, a majority of optimization models continue relying on classical gradient descent methods and face convergence issues such as slow convergence and being trapped in local minima and instabilities [39].

According to the review of literature, it can be seen that remarkable advancements have been made in the area of segmentation and classification of skin lesions using techniques like CNN, transformer networks, hybrid learning approaches, and optimization-based approaches. Despite these developments, many key challenges still exist, such as inaccurate boundary preservation, inadequate utilization of local and global features, poor optimization performance, and poor adaptability to noise and irregular dermoscopic environments [40]. Most of the available frameworks are not explicitly using an effective swarm-driven optimization strategy that can lead to better convergence of segments, and do not take advantage of segmentation context during the classification process. Thus, a comprehensive framework, which would improve lesion boundary detection, optimization, contextual feature extraction and adaptive feature fusion, is required. To overcome these challenges, an Adaptive Swarm-Driven Skin Lesion Segmentation and Classification Framework is suggested, which combines Adaptive Boundary-Aware U-Net (ABAU-Net), Hybrid Swarm Optimization (HSO), and Hybrid Vision Transformers with Dynamic Feature Fusion (HViT-DFF).

### III. Proposed Work

The proposed framework has two major stages; segmentation of skin lesions and classification of skin lesions. It includes 5 modules: image preprocessing, boundary-aware segmentation, swarm-based optimization, transformer-based classification and dynamic feature fusion. This step involves two main tasks: image preprocessing and image enhancement. The image preprocessing and enhancement are two main tasks in this step. Fig. 1 represents the architecture of the proposed adaptive swarm-driven skin lesion segmentation and classification framework.



**Fig. 1. Architecture of the proposed Adaptive Swarm-Driven Skin Lesion Segmentation and Classification Framework, comprising image preprocessing, ABAU-Net lesion segmentation, BOA-EHO optimization, CNN-Transformer feature extraction, dynamic feature fusion, and multi-class classification.**

The preprocessing stage is performed to enhance the quality of dermoscopic images before segmentation and classification. Dermoscopic images often contain illumination variations, noise, low contrast, hairs, and ruler artifacts that can affect diagnostic performance. An adaptive preprocessing pipeline is therefore employed to improve image quality while preserving important lesion characteristics. Adaptive contrast enhancement increases the intensity difference between the lesion and surrounding healthy skin regions. This process improves the visibility of subtle lesion structures and enhances the overall image appearance. Median filtering is applied to suppress noise while maintaining important edges and fine lesion boundaries. Morphological operations are subsequently used to remove hair and ruler artifacts without significantly altering lesion structures. The resulting enhanced images provide cleaner and more informative inputs for accurate segmentation and reliable classification. First, the images are color normalized to eliminate illumination variation to attenuate the effect of illumination differences, and then the images are resized and color normalized before inputting them into the segmentation network. The dermoscopic image is put into as:

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

where  $I_i$  in Eq.(1) [1] denotes the input dermoscopic image,  $H$  represents image height,  $W$  represents image width, and  $C$  indicates the number of color channels.

The enhanced image after preprocessing is computed as:

$$I_{enh} = N \left( R \left( M \left( A(C(I_i)) \right) \right) \right) \quad (2)$$

where  $I_{enh}$  in Eq. (2) [2] denotes the enhanced image,  $C(\cdot)$  represents adaptive contrast enhancement,  $A(\cdot)$  denotes artifact removal,  $M(\cdot)$  indicates median filtering,  $R(\cdot)$  represents image resizing, and  $N(\cdot)$  denotes normalization. The contrast enhancement operation is formulated as:

$$I_c = \frac{I_i - I_{min}}{I_{max} - I_{min}} \quad (3)$$

where  $I_c$  in Eq. (3) [3] denotes the contrast-enhanced image,  $I_{min}$  represents minimum pixel intensity, and  $I_{max}$  denotes maximum pixel intensity. The median filtering process is represented as:

$$I_m(x, y) = median \{I_c(i, j)\} \quad (4)$$

where  $I_m(x, y)$  in Eq. (4) [4] represents filtered pixel intensity at location  $(x, y)$ , and  $(i, j)$  indicates neighboring pixels within the filtering window. The normalized image is computed as:

$$I_{norm} = \frac{I_m - \mu}{\sigma} \quad (5)$$

where  $I_{norm}$  in Eq. (5) [5] denotes the normalized image,  $\mu$  represents mean pixel intensity, and  $\sigma$  indicates standard deviation of image intensity values.

#### A. Adaptive Boundary-Aware U-Net (ABAU-Net) for Lesion Segmentation

In the second stage, the proposed Adaptive Boundary-Aware U-Net (ABAU-Net) is used for precise skin lesion segmentation. ABAU-Net is an adaptive variant of the conventional U-Net with an encoder-decoder structure, which uses adaptive boundary-aware learning to

**Table 1.** Distribution of ISIC 2019 skin lesion images across training, validation, and testing sets for seven lesion classes.

| Skin Lesion Class          | Train ing Images | Valida tion Images | Testin g Images | Total Images |
|----------------------------|------------------|--------------------|-----------------|--------------|
| Basal Cell Carcinoma       | 302              | 37                 | 37              | 376          |
| Dermatofibr oma            | 77               | 9                  | 9               | 95           |
| Melanoma                   | 352              | 43                 | 43              | 438          |
| Nevus                      | 287              | 35                 | 35              | 357          |
| Pigmented Benign Keratosis | 370              | 46                 | 46              | 462          |
| Squamous Cell Carcinoma    | 145              | 18                 | 18              | 181          |
| Vascular Lesion            | 113              | 13                 | 13              | 139          |
| Total                      | 1646             | 201                | 201             | 2048         |

enhance segmentation at irregular lesion boundaries. Convolution, batch normalization, activation and down-sampling operations are performed to extract hierarchical spatial and semantic features from the encoder. The decoder recovers the segmentation map by using up-sampling and skip connection, maintaining the fine spatial details. A Boundary Refinement Module (BRM) is placed between the encoder and decoder to further refine the boundaries, highlighting the edges of lesions and enhancing propagation of contour features. The final result is a binary lesion mask which successfully separates the lesion from the healthy skin and will be used as an input for the classification stage. The encoder feature extraction is defined as:

$$E_l = \text{ReLU} \left( \text{BN}(\text{Conv}(E_{l-1})) \right) \quad (6)$$

where  $E_l$  in Eq. (6) [7] denotes encoder feature maps at level  $l$ ,  $\text{Conv}(\cdot)$  represents convolution operation,  $\text{BN}(\cdot)$  denotes batch normalization, and  $\text{ReLU}(\cdot)$  indicates rectified linear activation. The boundary refinement feature representation is computed as:

$$B_l = \sigma(\text{Conv}(E_l)) \odot E_l \quad (7)$$

where  $B_l$  in Eq. (7) [8] denotes boundary-enhanced feature representation,  $\sigma(\cdot)$  represents sigmoid activation, and  $\odot$  indicates element-wise multiplication. The refined boundary-aware feature map is obtained as:

$$F_{brm} = E_l + B_l \quad (8)$$

where  $F_{brm}$  in Eq. (8) [9] represents the refined boundary-aware feature map. The decoder reconstruction process is represented as:

$$D_l = \text{ReLU} \left( \text{BN} \left( \text{Conv}(\text{Concat}(\text{Up}(D_{l+1}), F_{brm})) \right) \right) \quad (9)$$

where  $D_l$  in Eq. (9) [10] denotes decoder feature representation,  $\text{Up}(\cdot)$  represents up-sampling operation, and  $\text{Concat}(\cdot)$  indicates skip-connection-based feature concatenation. The final predicted lesion mask is computed as:

$$M_{pred} = \sigma(\text{Conv}(D_1)) \quad (10)$$

where  $M_{pred}$  in Eq. (10) [10] denotes the predicted lesion segmentation mask. The Dice coefficient used for segmentation optimization is:

$$\text{Dice} = \frac{2|M_{gt} \cap M_{pred}| + \varepsilon}{|M_{gt}| + |M_{pred}| + \varepsilon} \quad (11)$$

where  $M_{gt}$  in Eq. (11) [12] represents ground-truth lesion mask,  $M_{pred}$  denotes predicted lesion mask, and  $\varepsilon$  indicates smoothing constant to avoid division instability. The segmentation loss function is expressed as:

$$L_{seg} = L_{dice} + L_{bce} + L_{boundary} \quad (12)$$

where  $L_{seg}$  in Eq. (12) [13] denotes total segmentation loss,  $L_{dice}$  represents Dice loss,  $L_{bce}$  denotes binary cross-entropy loss, and  $L_{boundary}$  indicates boundary-aware loss. Dataset Distribution for Skin Lesion Classification is shown in Table 1.

## B. Hybrid Swarm Optimization for Segmentation Convergence

The Butterfly Optimization Algorithm (BOA) and Elephant Herding Optimization (EHO) are adopted to enhance the optimization of segmentation and training stability of the ABAU-Net training process. The clan-update mechanisms improve local exploitation and convergence stability (EHO), while fragrance-based search (BOA) is used for global exploration. The fusion of these attributes leads to a well-balanced exploration-exploitation paradigm which in turn leads to faster convergence, stable optimization, better lesion boundary refinement, and more accurate segmentation of lesions. The optimization objective is formulated as:

$$\theta^* = \arg \min L_{seg}(\theta) \quad (13)$$

where  $\theta$  in Eq. (13) [14] represents trainable network parameters and  $\theta^*$  denotes optimal parameter set minimizing segmentation loss. The butterfly fragrance function is computed as:

$$f_i = c I_i^a \quad (14)$$

where  $f_i$  in Eq. (14) [15] denotes butterfly fragrance,  $c$  represents sensory modality coefficient,  $I_i$  indicates stimulus intensity, and  $a$  denotes power exponent controlling fragrance absorption. The BOA global search update equation is:

$$x_i^{t+1} = x_i^t + (r^2 \times g_{best} - x_i^t) \times f_i \quad (15)$$

where  $x_i^t$  in Eq. (15) [16] represents current solution position,  $x_i^{t+1}$  denotes updated solution position,  $r$  indicates random number between 0 and 1, and  $g_{best}$  represents global best solution.

The Elephant Herding Optimization clan update is defined as:

$$x_{new,c,i} = x_{c,i} + \alpha(x_{best,c} - x_{c,i}) \quad (16)$$

where  $x_{new,c,i}$  in Eq. (16) [17] denotes updated elephant position,  $x_{c,i}$  represents current elephant position in clan  $c$ ,  $\alpha$  denotes clan influence factor, and  $x_{best,c}$  indicates clan leader position. The hybrid swarm optimization update is expressed as:

$$\theta_{t+1} = \beta\theta_{BOA} + (1 - \beta)\theta_{EHO} \quad (17)$$

where  $\theta_{BOA}$  in Eq. (17) [18] represents Butterfly Optimization parameter update,  $\theta_{EHO}$  denotes Elephant Herding Optimization update, and  $\beta$  controls adaptive balance between exploration and exploitation.

### C. Hybrid Vision Transformer with Dynamic Feature Fusion (HVIT-DFF)

Once the lesion regions are segmented, the lesion regions are classified by the proposed Hybrid Vision Transformer with Dynamic Feature Fusion (HVIT-DFF). The model is a fusion of CNN based local feature extraction and transformer based global contextual learning to learn complementary lesion features. The segmented lesion ROI is represented as:

$$I_{roi} = I_{enh} \odot M_{pred} \quad (18)$$

where  $I_{roi}$  in Eq. (18) [19] denotes segmented lesion region of interest and  $\odot$  represents element-wise masking operation. The CNN-based local feature extraction is computed as:

$$F_{cnn} = CNN(I_{roi}) \quad (19)$$

where  $F_{cnn}$  in Eq. (19) [20] denotes local spatial and texture feature representation extracted using convolutional neural networks. The lesion image is divided into patches:

$$P = \{p_1, p_2, p_3, \dots, p_n\} \quad (20)$$

where  $P$  in Eq. (20) [21] represents the set of image patches and  $p_n$  denotes the  $n$ th image patch. The transformer patch embedding is represented as:

$$Z_0 = [p_{1E}; p_{2E}; \dots; p_{nE}] + E_{pos} \quad (21)$$

where  $Z_0$  in Eq. (21) [22] denotes embedded transformer input sequence,  $E$  represents patch embedding matrix, and  $E_{pos}$  indicates positional embedding. transformer self-attention operation is computed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (22)$$

where  $Q$  in Eq. (22) [23] denotes query matrix,  $K$  represents key matrix,  $V$  indicates value matrix, and  $d$  denotes feature dimension. The transformer encoder output is:

$$Z_l = MSA(LN(Z_{l-1})) + Z_{l-1} \quad (23)$$

where  $Z_l$  in Eq. (23) [25] denotes transformer encoder output at layer  $l$ ,  $MSA(\cdot)$  represents multi-head self-attention, and  $LN(\cdot)$  denotes layer normalization. The final transformer feature representation is:

$$F_{vit} = Z_L \quad (24)$$

where  $F_{vit}$  in Eq. (24) [26] denotes transformer-based global contextual feature representation.

## IV. Results

The proposed framework has been tested on the ISIC 2019 dataset that consists of 2,048 dermoscopic images belonging to seven skin lesion categories. Preprocessing was performed on the images which included adaptive contrast enhancement, removal of the hair artifacts, filtering, color normalization, and resizing to  $224 \times 224 \times 3$  dimensions. Classification Performance of the Proposed HVIT-DFF Model is

**Table 2.** Classification performance comparison of the proposed HVIT-DFF model and baseline deep-learning models on the test set

| Method                 | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC   |
|------------------------|--------------|---------------|------------|--------------|-------|
| CNN                    | 91.8         | 91.2          | 91.0       | 91.1         | 0.931 |
| ResNet50               | 94.5         | 94.2          | 94.1       | 94.1         | 0.954 |
| DenseNet121            | 95.6         | 95.2          | 95.3       | 95.2         | 0.963 |
| Vision Transformer     | 96.8         | 96.5          | 96.3       | 96.4         | 0.974 |
| CNN-Transformer Hybrid | 97.5         | 97.2          | 97.1       | 97.1         | 0.981 |
| Proposed HVIT-DFF      | 98.8         | 98.6          | 98.5       | 98.5         | 0.992 |

**Table 3.** Five-fold cross-validation performance of the proposed ABAU-Net/HVIT-DFF framework on the ISIC 2019 dataset.

| Fold    | Dice (%) | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---------|----------|--------------|---------------|------------|--------------|
| Fold-1  | 97.9     | 98.4         | 98.2          | 98.1       | 98.1         |
| Fold-2  | 98.1     | 98.6         | 98.4          | 98.3       | 98.3         |
| Fold-3  | 98.3     | 98.9         | 98.7          | 98.5       | 98.5         |
| Fold-4  | 98.2     | 98.8         | 98.5          | 98.4       | 98.4         |
| Fold-5  | 98.4     | 99.0         | 98.8          | 98.7       | 98.7         |
| Average | 98.2     | 98.8         | 98.5          | 98.4       | 98.4         |

represented in Table 2. The data set consisted of 1,646 training images, 201 validation images and 201 testing images, using five-fold cross-validation. The five-fold cross validation also verified the stability of learning and good generalization ability of the model, with average Dice, Accuracy, Precision, Recall, and F1-score of 98.2%, 98.8%, 98.5%, 98.4%, and 98.4%, respectively shown in Table 3. Fig. 2 shows the qualitative segmented image of the proposed work.

### A. Dynamic Feature Fusion and Multi-Class Lesion Classification

The proposed Dynamic Feature Fusion (DFF) module

The proposed Adaptive Swarm-Driven Skin Lesion Segmentation and Classification Framework consists of two phases: (1) skin lesion segmentation and (2) multi-class skin lesion classification. Adaptive contrast enhancement, hair artifact removal, noise reduction, color normalization and resizing are performed as the first step of dermoscopic image pre-processing. The improved images subsequently go through the proposed Adaptive Boundary-Aware U-Net (ABAU-Net) along with the Boundary Refinement Module (BRM), where the latter is used to segment the lesions and preserve the boundaries accurately. A hybrid Butterfly Optimization Algorithm–Elephant Herding

| Case Type                             | Class (Abbr.)                    | Original Image | Ground Truth Mask | Predicted Mask (ABAU-Net) | Overlay (GT in Green, Pred in Red) | Boundary Zoom (Overlay)       |
|---------------------------------------|----------------------------------|----------------|-------------------|---------------------------|------------------------------------|-------------------------------|
| Easy Cases (Well-defined Lesions)     | Melanoma (MEL)                   |                |                   |                           |                                    | <br>Dice: 0.982<br>IoU: 0.965 |
|                                       | Nevus (NV)                       |                |                   |                           |                                    | <br>Dice: 0.974<br>IoU: 0.949 |
|                                       | Basal Cell Carcinoma (BCC)       |                |                   |                           |                                    | <br>Dice: 0.968<br>IoU: 0.941 |
| Difficult Cases (Challenging Lesions) | Melanoma (MEL)                   |                |                   |                           |                                    | <br>Dice: 0.912<br>IoU: 0.837 |
|                                       | Pigmented Benign Keratosis (PBK) |                |                   |                           |                                    | <br>Dice: 0.901<br>IoU: 0.807 |
|                                       | Basal Cell Carcinoma (BCC)       |                |                   |                           |                                    | <br>Dice: 0.894<br>IoU: 0.792 |

Fig. 2. Qualitative comparison of skin lesion segmentation results using the proposed ABAU-Net. Columns show the original dermoscopic image, ground-truth mask, predicted mask, ground-truth/prediction overlay, and enlarged boundary comparison for representative easy and challenging lesion cases.

adaptively fuses the features from the convolutional network and transformer network, with dynamic weights assigned to the features based on their context. These fused features are then passed through fully connected layers having the Softmax activation function, for multi-class classification. The adaptive fusion weights are computed as:

$$w_{cnn} = \frac{\exp(w_c F_{cnn})}{\exp(w_c F_{cnn}) + \exp(w_t F_{vit})} \quad (25)$$

$$w_{vit} = \frac{\exp(w_t F_{vit})}{\exp(w_c F_{cnn}) + \exp(w_t F_{vit})} \quad (26)$$

where  $w_{cnn}$  and  $w_{vit}$  in Eq. (25) and Eq. (26) [27] denote adaptive fusion weights for CNN and transformer.

Optimization (BOA-EHO) strategy is applied to the network parameters optimization and better convergence during training. Ablation study of the proposed work is shown in Table 4. The proposed Hybrid Vision Transformer with Dynamic Feature Fusion (HVIT-DFF) is applied to segment the lesion regions, segment local feature representations by CNN, and segment global contextual information by Transformer, and then the Dynamic Feature Fusion (DFF) module combines the two feature representations in an adaptive way to accurately classify the lesion into multiple categories by a Softmax classifier. The segmentation performance is assessed

via Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision and Recall as defined in equations Eq. (27)–Eq. (30) [30]. The DSC in Eq. (27) measures the overlap accuracy between the predicted lesion mask and the ground-truth segmentation by considering true positive (TP), false positive (FP), and false negative (FN) pixels, where higher DSC values indicate better segmentation performance. The IoU metric in Eq. (28) evaluates the ratio between the intersection and union of predicted and ground-truth lesion regions, thereby quantifying segmentation consistency. Precision in Eq. (29) measures the proportion of correctly predicted lesion pixels among all predicted positive pixels, reflecting the model's ability to minimize false positive detections. Recall in Eq. (30) evaluates the proportion of actual lesion pixels correctly identified by the model, indicating the effectiveness of lesion region detection while minimizing false negatives. In this study, 2,048 dermoscopic images of different lesion classes (ISIC 2019) were used. Duplicate, corrupted, poorly-illuminated, and poorly-annotated images were omitted. Seven clinically relevant lesion classes were considered, namely Basal Cell Carcinoma (376 images), Dermatofibroma (95 images), Melanoma (438 images), Nevus (357 images), Pigmented Benign Keratosis (462 images), Squamous Cell Carcinoma (181 images), and Vascular Lesion (139 images). The images were preprocessed with adaptive contrast enhancement, hair artifact removal, median filtering, color normalization, intensity normalization and resized to  $224 \times 224 \times 3$ . The data was augmented online by random flipping, random rotation  $\pm 20^\circ$ , random scaling 0.8–1.2, random cropping, brightness/contrast adjustments and Gaussian noise. Patients were stratified into training (1646 images), validation (201 images), and testing (201 images) sets for the dataset; five-fold cross-validation was used to assess robustness. We have designed and tested the proposed models, ABAU-Net and HViT-DFF, in Python 3.11 with PyTorch 2.2 and trained with the Adam optimizer, an initial learning rate of  $1 \times 10^{-4}$ , weight decay of  $1 \times 10^{-5}$ , batch size of 16, 100 training epochs, a cosine annealing learning-rate scheduler, dropout rate of 0.30, and a fixed random seed of 42. Global exploration and local exploitation were balanced in the hybrid BOA-EHO algorithm by incorporating it with backpropagation, leading to better convergence, boundary preservation and segmentation accuracy. Experiments on an NVIDIA RTX 4090 GPU proved the effectiveness and robustness of the proposed framework. For optimization, the hybrid BOA-EHO algorithm utilized a sensory modality coefficient ( $c = 0.01$ ), power exponent ( $a = 0.10$ ), switching probability ( $p = 0.80$ ), clan influence factor ( $\alpha = 0.50$ ), separating factor ( $\beta_{\text{EHO}} = 0.10$ ), and 5 clans, with a swarm population of 30 candidate solutions and 100 optimization iterations. An

adaptive exploration–exploitation coefficient ( $\beta = 0.50$ ) was dynamically updated throughout training. The BOA-EHO optimization operated in conjunction with gradient-based backpropagation.

$$DSC = \frac{2TP}{(2TP + FP + FN)} \quad (27)$$

$$IoU = \frac{TP}{(TP + FP + FN)} \quad (28)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (29)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (30)$$

This framework was implemented in Python 3.11, with PyTorch 2.2, and trained on an NVIDIA RTX 4090 GPU, using data augmentation and Adam optimization. The proposed ABAU-Net gives a Dice coefficient of 98.2%, Jaccard index of 97.5%, Precision of 98.0%, Recall of 98.1% and Accuracy of 97.2%. The HViT-DFF classifier gave the highest accuracy of 98.8%, precision of 98.6%, and recall of 98.5%, F1 score of 98.5%, and AUC of 0.992%, among all CNN, Transformer and hybrid models.

**Table 4. Ablation study of the proposed framework showing the incremental effects of preprocessing, BRM, BOA, EHO, CNN features, Transformer features, and dynamic feature fusion**

| Configuration               | Dice (%)    | Jaccard (%) | Accuracy (%) | AUC          |
|-----------------------------|-------------|-------------|--------------|--------------|
| Baseline CNN–Transformer    | 95.8        | 94.2        | 95.6         | 0.972        |
| + Preprocessing             | 96.4        | 95.0        | 96.2         | 0.978        |
| + BRM                       | 97.1        | 96.1        | 96.9         | 0.982        |
| + BOA                       | 97.4        | 96.5        | 97.3         | 0.985        |
| + EHO                       | 97.6        | 96.8        | 97.6         | 0.987        |
| + CNN Features              | 97.8        | 97.0        | 97.9         | 0.988        |
| + Transformer Features      | 98.0        | 97.2        | 98.2         | 0.990        |
| + DFF (Proposed Full Model) | <b>98.2</b> | <b>97.5</b> | <b>98.8</b>  | <b>0.992</b> |

## B. Performance Analysis and Statistical Validation

The proposed framework achieved an AUC of 0.992, classification accuracy of 98.8%, Dice Coefficient of 98.2% and Jaccard Index of 97.5% on ISIC dataset, which shows excellent segmentation and classification performance. The model exhibited high stability in learning and good generalization, as demonstrated by the average Dice, accuracy, precision, recall, and F1-score of  $98.2 \pm 0.19\%$ ,  $98.8 \pm 0.24\%$ ,  $98.5 \pm 0.24\%$ ,

**Table 5.** Quantitative comparison of the proposed ABAU-Net + HViT-DFF framework with recent skin lesion segmentation and classification methods on the ISIC benchmark dataset

| Author              | Year | Method                               | Limitation                                                                                     | Output (Dice / Jaccard / Accuracy / AUC) |
|---------------------|------|--------------------------------------|------------------------------------------------------------------------------------------------|------------------------------------------|
| Amin et al. [12]    | 2025 | Boundary-Aware CNN–Transformer       | Boundary attention alone struggles with hair/illumination artifacts; no adaptive preprocessing | 97.1% / 96% / 97.3% / 0.98               |
| Fiaz et al. [10]    | 2025 | Explainable Hybrid Framework         | Interpretability constraints limit representational capacity, lowering raw performance         | 96.8% / 95% / 96.9% / 0.97               |
| Azhari et al. [21]  | 2025 | Hybrid Transformer–CNN               | Standard optimizer limits convergence compared to swarm-based tuning                           | 97.4% / 96% / 97.8% / 0.98               |
| Naeem et al. [23]   | 2026 | LEDNet + Swin-UMamba                 | State-space backbone generalizes less well across decision thresholds (lower AUC)              | 97.2% / 96% / 97.1% / 0.97               |
| Binzagr et al. [13] | 2026 | SAM Adapter + Vision Transformer     | Lacks domain-specific optimization despite strong general segmentation transfer                | 97.6% / 96% / 97.9% / 0.98               |
| Proposed            | —    | <b>ABAU-Net + HViT-DFF (BOA-EHO)</b> | Higher computational complexity; single-dataset validation                                     | <b>98.2% / 97% / 98.8% / 0.99</b>        |

98.4 ± 0.22%, and 98.4 ± 0.22%, respectively, when evaluated using five-fold cross-validation. The proposed HViT-DFF correctly classified about 199 images, and the accuracy rate was 98.8% in the independent test set consisting of 201 images with only two images misclassified. The statistical significance analysis also confirmed the validity of the proposed framework with the p-value of the paired t-test and McNemar’s test models being less than 0.05 for CNN, ResNet50, DenseNet121, Vision Transformer, and CNN-Transformer Hybrid models.

V. Discussion

The proposed framework’s results reflect strong performance at both pixel and case level. The Dice Coefficient of 98.2% and Jaccard Index of 97.5% show that ABAU-Net’s predicted lesion boundaries closely match expert annotations, with only small, localized errors rather than systematic misses. Precision of 98.0% indicates that when the model labels a pixel or lesion as positive, meaning lesion tissue or a malignant class, it is correct 98% of the time, implying a low false-positive rate. In a screening context, this translates to fewer healthy patients being incorrectly flagged for

further invasive testing. Recall, or sensitivity, of 98.1% shows that the model correctly identifies 98.1% of actual lesions or malignant cases, meaning very few true melanomas or lesion regions are missed. In dermatological screening, this metric is arguably the most clinically sensitive one, since a missed melanoma carries far greater risk than a false alarm. Classification Accuracy of 98.8% confirms that HViT-DFF’s fusion of CNN-based local features and transformer-based global context effectively distinguishes multiple lesion classes, and the AUC of 0.992 shows strong discriminative ability across decision thresholds. Together, these metrics suggest the BOA-EHO optimizer improved both segmentation and classification simultaneously, rather than trading one off against the other. As shown in Table 5, proposed method consistently outperforms recent CNN-, Transformer-, and hybrid-based methods, though by modest, incremental margins (0.6–1.4 points in Dice; 0.9–1.9 points in accuracy). The closest competitors, Azhari et al. (2025) [21] and Binzagr et al. (2026) [13], reflect the field’s shift toward hybrid Transformer–CNN and foundation-model-adapted architectures, but neither incorporates adaptive preprocessing or swarm-

based optimization, which likely accounts for its edge. Fiaz et al. (2025) [10] prioritizes explainability at some cost to raw performance, highlighting a trade-off does not address. Given the statistical significance reported ( $p < 0.05$ ), these gains appear genuine rather than incidental, though they remain incremental given how mature the ISIC benchmark has become. The framework was evaluated only on ISIC 2019, limiting generalizability to other populations, devices, and skin tones. Its dual-network design with BOA-EHO optimization adds computational cost, restricting use on resource-constrained devices. Table 6 shows the similarity and difference analysis of existing models. These results carry several implications for computer-aided dermatology. The near-ceiling Dice, Jaccard, and AUC values suggest frameworks could serve as reliable diagnostic support tools, extending the boundary-preserving approach introduced by Amin et al. (2025) [12] with adaptive preprocessing that further

improves robustness to real-world image artifacts. The high sensitivity achieved also builds on the direction set by Azhari et al. (2025) [12] and Binzagr et al. (2026) [13], whose hybrid Transformer-CNN and SAM-adapted architectures demonstrated the value of combining local and global feature representations, suggesting that such fusion strategies are increasingly viable for large-scale screening and teledermatology applications where minimizing missed lesions is critical. The comparison with Fiaz et al. (2025) [10] implies that performance gains from swarm-based optimization need not come at the cost of interpretability being addressed altogether; rather, future integration of explainability mechanisms alongside BOA-EHO-style optimization could combine the strengths of both approaches. Similarly, the efficiency-oriented design of Naeem et al. (2026) [23], using a lightweight Swin-UMamba backbone, points toward a practical implication for ASSLA: that future work should adapt

Table 6. Similarity and Difference Analysis of Existing Models

| Comparison                                               | Similarity                                                                                                                                                 | Key Difference                                                                                                                                                             |
|----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Amin et al. (2025) [12] vs. Azhari et al. (2025) [21]    | Both employ deep-learning-based approaches for accurate skin-lesion segmentation and exploit learned visual representations to improve lesion delineation. | Amin et al. emphasize boundary-aware segmentation, whereas Azhari et al. employ a hybrid CNN-Transformer architecture to combine local and global feature representations. |
| Amin et al. (2025) [12] vs. Fiaz et al. (2025) [10]      | Both aim to improve automated skin-lesion analysis using deep-learning techniques.                                                                         | Amin et al. primarily concentrate on accurate lesion-boundary preservation, while Fiaz et al. place greater emphasis on explainability and interpretable analysis.         |
| Azhari et al. (2025) [21] vs. Fiaz et al. (2025) [10]    | Both employ deep-learning models for automated skin-lesion analysis and classification/segmentation tasks.                                                 | Azhari et al. focus on hybrid CNN-Transformer feature representation, whereas Fiaz et al. emphasize explainability and interpretable predictions.                          |
| Azhari et al. (2025) [21] vs. Binzagr et al. (2026) [13] | Both exploit advanced feature representations combining complementary lesion information to improve automated lesion analysis.                             | Azhari et al. explicitly use a CNN-Transformer hybrid architecture, whereas Binzagr et al. employ SAM adaptation for lesion segmentation and analysis.                     |
| Azhari et al. (2025) [21] vs. Naeem et al. (2026) [23]   | Both use advanced neural architectures to capture complex visual characteristics of skin lesions.                                                          | Azhari et al. emphasize hybrid CNN-Transformer representation, whereas Naeem et al. use Swin-UMamba with an emphasis on lightweight and efficient segmentation.            |
| Fiaz et al. (2025) [10] vs. Binzagr et al. (2026) [13]   | Both apply deep-learning techniques to automated skin-lesion analysis and seek reliable computer-aided diagnostic support.                                 | Fiaz et al. emphasize model explainability, whereas Binzagr et al. focus on SAM-based lesion representation and segmentation.                                              |
| Fiaz et al. (2025) [10] vs. Naeem et al. (2026) [23]     | Both address automated skin-lesion analysis using deep neural networks.                                                                                    | Fiaz et al. prioritize interpretability, whereas Naeem et al. prioritize computational efficiency through a lightweight Swin-UMamba architecture.                          |
| Binzagr et al. (2026) [13] vs. Naeem et al. (2026) [23]  | Both employ advanced architectures for accurate automated lesion segmentation and representation.                                                          | Binzagr et al. utilize SAM adaptation, whereas Naeem et al. employ Swin-UMamba with a lightweight design aimed at reducing computational requirements.                     |

swarm-optimized fusion strategies into more computationally efficient backbones to enable point-of-care and mobile deployment, rather than relying solely on the heavier dual-network structure used in this study. Table 5 presents a quantitative comparison between the proposed Adaptive Swarm-Driven Skin Lesion Analysis Framework and recently published skin lesion segmentation and classification approaches evaluated on ISIC benchmark datasets.

## V. Conclusion

This study proposed the Adaptive Swarm-Driven Skin Lesion Analysis (ASSLA) framework, which effectively combines ABAU-Net for boundary-aware segmentation, BOA-EHO optimization for improved convergence, and HViT-DFF for multi-class classification through the fusion of local and global feature representations. Evaluated on the ISIC 2019 dataset, the framework achieved a Dice Coefficient of 98.2%, Jaccard Index of 97.5%, Precision of 98.0%, Recall of 98.1%, Classification Accuracy of 98.8%, and AUC of 0.992, consistently outperforming recent CNN-, Transformer-, and hybrid-based methods with statistically significant improvements ( $p < 0.05$ ). These results confirm that integrating adaptive preprocessing, boundary-preserving segmentation, swarm-based optimization, and dynamic feature fusion yields measurable gains in both segmentation precision and classification reliability. The framework's added computational complexity and its evaluation within the scope of a single dataset temper the immediate clinical readiness of these results. Consequently, future work should prioritize multi-center clinical validation to establish generalizability across diverse populations, imaging devices, and skin tones, alongside efforts to reduce computational overhead through lightweight and efficient architectural variants suitable for point-of-care and mobile deployment. Further extension of the framework toward multimodal learning, incorporating clinical metadata or complementary imaging modalities, and toward self-supervised learning techniques that reduce dependence on large labeled datasets.

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Data Availability:** The dataset used in this study is publicly available from the ISIC/Kaggle repository at [URL]. No new clinical dataset was generated during this study.

**Code Availability:** The code used in this study is available from the corresponding author upon reasonable request.

**Author Contributions:** Thilagavathi Rajamanickam: Conceptualization, Methodology, Writing Original Draft.

Anuradha Konidena: Data Curation, Software, Investigation. N. Mahesh: Software, Validation, Formal Analysis. M. Balakrishnan: Formal Analysis, Visualization, Writing, Review & Editing. N.K. Priyadharsini: Supervision, Validation, Writing, Review & Editing. Suganthi Rajamanickam: Methodology, Investigation, Formal Analysis, Writing, Review & Editing. All authors reviewed and approved the final manuscript.

**Ethical Approval:** Ethical approval was not required because the study used only publicly available, de-identified secondary image data and involved no direct recruitment or intervention involving human participants.

**Consent to Participate:** Not applicable.

**Consent for Publication:** Not applicable because no identifiable individual participant information collected directly by the authors is reported.

**Competing Interests:** The authors declare that they have no competing financial or non-financial interests.

**Generative AI and AI-Assisted Technologies:** The authors declare that no generative artificial intelligence or AI-assisted technologies were used in the writing, editing, analysis, or preparation of this manuscript.

**Dataset Use Statement:** The dataset was used in accordance with the licensing and terms of use specified by the original repository.

**Acknowledgments:** Any individuals or institutions providing non-author technical, administrative, or material support should be acknowledged.

## References

- [1] Hamza, A., & Damaševičius, R. "Deep learning for brain tumor segmentation and classification: a systematic review of methods and trends", *Computers, materials and continua.*, vol.86(1), pp.1-41, 2025, DOI:10.32604/cmc.2025.069721.
- [2] Debelee, T. G. "Skin lesion classification and detection using machine learning techniques: a systematic review", *Diagnostics*, vol.13(19), pp.3147, 2023, DOI:10.3390/diagnostics13193147.
- [3] Al-Masni, M. A., Kim, D. H., & Kim, T. S. "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification", *Computer methods and programs in biomedicine*, vol.190, Art. No.105351, 2020, DOI:10.1016/j.cmpb.2020.105351.
- [4] Mahbod, A., Tschandl, P., Langs, G., Ecker, R., & Ellinger, I. "The effects of skin lesion segmentation on the performance of dermoscopic image

- classification", Computer Methods and Programs in Biomedicine, vol.197, Art.No.105725, 2020, DOI:10.1016/j.cmpb.2020.105725.
- [5] Khalaf, A. D., Hamdan, H., Halin, A. B. A., & Manshor, N. "Segmentation and classification of skin cancer diseases based on deep learning: Challenges and future directions", IEEE Access, vol.31,2025,DOI:10.1109/ACCESS.2025.3569170
- [6] R. Kasmi, K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic abcd rule", IET Image Proc. Vol.10 (6) 2016, pp.448–455, DOI:10.1049/iet-ipr.2015.0385
- [7] Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. "Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey", *Wireless Pers Commun* vol.137, pp.1685–1714, 2024, DOI:10.1007/s11277-024-11428-1
- [8] Mehdar, K. M., Soomro, T. A., Ali, A., Bin Ubaid, F., Irfan, M., Halawani, H. T., ... & Abdelkafi Magzoub, M. "Deep neural network-based robust framework for automated skin lesion segmentation and analysis", Digital Health, vol.12, 2026, Art.No.20552076261427501, DOI:10.1177/20552076261427501
- [9] Gabani, V., Navamani, T. M., Shyamala, K., & Vaswani Rajpal, V. K. "Multimodal skin lesion classification for early cancer diagnosis using deep learning", Frontiers in Physiology, vol.17, Art. No.1717517, DOI:10.3389/fphys.2026.1717517
- [10] Fiaz, M., Shoaib Khan, M. B., Khan, A. H., Bilal, A., Abdullah, M., Darem, A. A., & Sarwar, R. "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification", Frontiers in Medicine, vol.12, Art. No.1681542, DOI:10.3389/fmed.2025.1681542
- [11] Attique Khan M, Sharif M, Akram T, Kadry S, Hsu CH, "A two-stream deep neural network-based intelligent system for complex skin cancer types classification", Int J Intell Syst vol.37(12), pp.10621-10649, 2022, <https://doi.org/10.1002/int.22691>
- [12] Amin, J., Azhar, M., Arshad, H., Zafar, A., & Kim, S. H., "Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks", Frontiers in Medicine, vol.12, Art. No.1524146, DOI:10.3389/fmed.2025.1524146
- [13] Binzagr, F., & Hariri, M. "Foundation-Model-Driven Skin Lesion Segmentation and Classification Using SAM-Adapters and Vision Transformers", Diagnostics, vol.16(3), DOI:10.3390/diagnostics16030468
- [14] Lilhore, U. K., Anitha, D., Priya, R., Rahane, W. P., Bangare, S. L., Kavitha, R., ... & Simaiya, S. "Adaptive Hybrid AI Framework for Robust and Explainable Skin Lesion Segmentation and Melanoma Detection", International Journal of Computational Intelligence Systems, vol. 19(1), 2026, DOI:10.1007/s44196-026-01233-y
- [15] Huang, D.X., Zhou, X.H., Xie, X.L., Liu, S.Q., Wang, S., Feng, Z.Q., Gui, M.J., Li, H., Xiang, T.Y., Yao, B.X., et al., "Spironet: Spatialfrequency learning and graph-based channel interaction network for vessel segmentation", Biomedical Signal Processing and Control, vol. 126, Art. No.110814, 2026, DOI: 10.1016/j.bspc.2026.110814
- [16] Naeem, M. A., Yang, S., Saleem, M. A., Javeed, A., & Ahmad, T., "A hybrid approach for accurate skin lesion segmentation using LEDNet and Swin-UMamba", Scientific Reports, vol.16(1), pp.5415, 2026, DOI:10.1038/s41598-026-38056-y
- [17] Verma, N., Ranvijay, R., & Yadav, D. K., "Robust Segmentation of Skin Lesions via Confidence-Guided ConvLSTM Integrated With Advanced Curriculum Strategies", International Journal of Imaging Systems and Technology, vol.36(2), pp.e70307,2026, DOI:/10.1002/ima.70307
- [18] Khan, M. A., Sharif, M. I., Raza, M., Anjum, A., Saba, T., & Shad, S. A., "Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection", Expert Systems, vol.39(7), pp. e12497, DOI:10.1111/exsy.12497
- [19] Qamar, S., Alkhatarishi, M., Alam, F., & Fazil, M., "Confidence-weighted semi-supervised learning for skin lesion segmentation using hybrid CNN-Transformer networks", IEEE Access, vol.14, pp.24579–24594, 2026, DOI:10.1109/ACCESS.2026.3663774
- [20] Ali, G., Awang, M. K., Rashid, J., Hamza, M., Khashan, O. A., & Ghani, A., "WIDENet: A Novel Lightweight CNN for Robust Skin Lesion Segmentation", IEEE Access, vol.14, pp.57763 – 57781, 2026, DOI:10.1109/ACCESS.2026.3683571
- [21] Azhari, A. A., Yudistira, N., Widodo, A. W., & Yagi, Y., "Two-stage CNN with weakly supervised segmentation for skin lesion classification", Multimedia Tools and Applications, vol.84(41), pp.49769-49800, 2025, DOI:/10.1007/s11042-025-21091-8
- [22] Lakshmi, R., & Arthi, B., "Superior skin cancer segmentation and classification via vision transformer-based ResNet with long short-term memory", Communications in Statistics-Simulation and Computation, pp.1-30, 2026, DOI:10.1080/03610918.2026.2647955
- [23] Naeem, S., Haneef, F., & Noor, M. N., "A hybrid deep learning based robust framework for enhancing the prediction of skin lesions", Network Modeling Analysis in Health Informatics and Bioinformatics, vol.15(1), 2026, DOI:10.1007/s13721-025-00714-y

- [24] N. Moradi, N. Mahdavi-Amiri, "Kernel sparse representation-based model for skin lesions segmentation and classification", *Comput. Methods Programs Biomed.*, vol. 182, Art. No.105038, DOI:10.1016/j.cmpb.2019.105038
- [25] R.F. Mansour, S.A. Althubiti, F. Alenezi, "Computer vision with machine learning enabled skin lesion classification model", *Computers, Materials & Continua*, vol. 73(1), pp.849-864, 2022, DOI:10.32604/cmc.2022.029265
- [26] V. Srividhya, K. Sujatha, R.S. Ponmagal, G. Durgadevi, L. Madheshwaran, "Visionbased detection and categorization of skin lesions using deep learning neural networks", *Procedia Comput. Sci.*, vol.171, pp.1726–1735, 2020, DOI:10.1016/j.procs.2020.04.185
- [27] M. Rastgoo, R. Garcia, O. Morel, F. Marzani, "Automatic differentiation of melanoma from dysplastic nevi", *Comput. Med. Imaging Graph.*, vol.43, pp.44–52, 2015, DOI:10.1016/j.compmedimag.2015.02.011
- [28] H. Zhao, G. Wang, Y. Wu, H. Wang, Y. Li, "Vcmix-net: a hybrid network for medical image segmentation", *Biomed. Signal Process. Control*, vol. 86, Art. No.105241, 2023, DOI:10.1016/j.bspc.2023.105241
- [29] S. Jain, N. Pise, "Computer-aided melanoma skin cancer detection using image processing", *Procedia Comput. Sci.*, vol. 48, pp. 735–740, 2015, DOI:10.1016/j.procs.2015.04.209
- [30] R. Kasmi, K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic abcd rule", *IET Image Proc.*, vol. 10 (6), pp.448–455, 2015, DOI:10.1049/iet-ipr.2015.0385
- [31] G. Yang, S. Luo, P. Greer, "Advancements in skin cancer classification: a review of machine learning techniques in clinical image analysis", *Multimed. Tools Appl.*, vol. 84, pp. 9837–9864, 2025, DOI:10.1007/s11042-024-19298-2
- [32] M. Zafar, M.I. Sharif, M.I. Sharif, S. Kadry, S.A.C. Bukhari, H.T. Rauf, "Skin lesion analysis and cancer detection based on machine/deep learning techniques: a comprehensive survey", *Life*, vol. 13 (1), 2023, DOI:10.3390/life13010146
- [33] Ioannis Giotis, Niki Molders, Steffen Land, Michael Biehl, Marcel F Jonkman, and Nicolai Petkov., "Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images", *Expert Systems with Applications*, vol.42(19), pp.6578–6585, 2015, DOI:10.1016/j.eswa.2015.04.034
- [34] S. S. Mohammed and J. M. Al-Tuwaijari., "Skin disease classification system based on machine learning technique: A survey", In *IOP Conference Series: Materials Science and Engineering*, vol. 1076, pp.012045. 2021, DOI:10.1088/1757-899X/1076/1/012045
- [35] M. Dildar, S. Akram, M. Irfan, H.U. Khan, M. Ramzan, A.R. Mahmood, M. H. Mahnashi, "Skin cancer detection: a review using deep learning techniques", *Int. J. Environ. Res. Public Health*, vol. 18 (10), pp. 5479, 2021, DOI:10.3390/ijerph18105479
- [36] T. Goswami, V. K. Dabhi, and H. B. Prajapati, "Skin disease classification from image-a survey", In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 599–605. IEEE, 2020, DOI:10.1109/ICACCS48705.2020.9074232
- [37] E. Vocaturo, E. Zumpano, and P. Veltri. "Image pre-processing in computer vision systems for melanoma detection", In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2117–2124. IEEE, 2018, DOI:10.1109/BIBM.2018.8621507
- [38] S. Pathan, K.G. Prabhu, P.C. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review", *Biomed. Signal Process. Control*, vol.39, pp.237–262, 2018, DOI: 10.1016/j.bspc.2017.07.010
- [39] Z. Huang, H. Deng, S. Yin, T. Zhang, W. Tang, Q. Wang, "Adf-net: a novel adaptive dual-stream encoding and focal attention decoding network for skin lesion segmentation", *Biomed. Signal Process. Control*, vol.91, Art. No.105895, 2024, DOI:10.1016/j.bspc.2023.105895
- [40] R.R. Babu, F.M. Philip, "Optimized deep learning for skin lesion segmentation and skin cancer detection", *Biomed. Signal Process. Control*, vol. 95, Art. No. 106292, 2024, DOI:10.1016/j.bspc.2024.106292

### Author Biography



**Dr. Thilagavathi Rajamanickam** is an Assistant Professor in the Department of Mathematics at Rajalakshmi Engineering College, Chennai. She has 12 years of teaching experience and 3 years of full-time research experience. She earned her B.Sc. in Mathematics from Seethalakshmi Ramaswamy College (BDU, Trichy), her M.Sc. in Mathematics from Bharathidasan University (BDU, Trichy), her M.Phil. in Mathematics from Dhanalakshmi Srinivasan Arts and Science College (BDU, Trichy), and her P.h.D. from Vel Tech Rangarajan Dr.Sagunthala R & D Institute of Science & Technology, Chennai. To date, she has published 6 research papers in international journals and

conference proceedings. Her core research areas include two-warehouse inventory, optimization techniques, supply chain management, and operations research.



**Dr. K. Anuradha Konidena**

is currently serving as an Associate Professor in the Department of Computer Science and Engineering at KL University. She has an illustrious career spanning over 23 years in academics, research, and administration, with a strong commitment to advancing the fields of Artificial Intelligence and Data Science. As a dedicated academician, she has significantly contributed to curriculum development, student mentoring, and promoting research excellence. Her expertise, leadership, and passion for innovation have inspired both students and faculty members, while her active involvement in academic and research initiatives continues to strengthen the growth of emerging technologies in higher education.



**Dr. N. Mahesh** is serving as an Associate Professor in the Department of Electronics and Instrumentation Engineering at . He completed his Ph.D. from in 2020 and has been associated with Kongu Engineering College since 2009. His areas of

specialization include Embedded Systems, Instrumentation, and Control Systems. Dr. N. Mahesh has contributed significantly to academic and research activities through journal publications, conference papers, and student mentoring. He has published several research articles in reputed journals and conferences, earning notable citations in his field of study. He is also associated with professional research platforms such as ORCID, Scopus, and Google Scholar, reflecting his active involvement in research and technical development.

**Dr. M. Balakrishnan** is a Professor in the Department of Artificial Intelligence and Data Science at Dr. Mahalingam College of Engineering and Technology. He earned his Doctorate from Anna University in 2014 and completed his M.E. in Computer Science and Engineering from K.S.Rangasamy College of



Technology in 2006. With nearly 19 years of teaching and academic experience, he has contributed significantly to the fields of Artificial Intelligence, Data Science, and Computer Science education. His areas of interest include Artificial Intelligence, Machine Learning, Data Analytics, and emerging technologies. He is actively involved in teaching, research, academic administration, and student mentoring activities.



**Dr. N. K. Priyadharsini**

is presently working as Associate Professor in Department of Computer Science and Engineering, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu. She has 15 years of teaching and industry experience. She completed PhD degree in the area of Video processing. Her major areas of interest include Data Structures, Operating Systems, Design and Analysis of Algorithms, Image Processing and Web applications. She holds membership in various professional bodies like ISTE, IAETSD, IAENG and IRED. She has published 15 papers in National and International Conferences and Journals. She has attended more than 45 National Seminars/Workshops and instructor for CCNA Courses. She is a Microsoft Certified Professional for developing Web Applications and Designing and Implementing Database with Microsoft SQL server.



**Dr Suganthi Rajamanickam,**

B.E., M.E., Ph.D., is currently serving as an Associate Professor in the Department of Electronics and Communication Engineering at Panimalar Engineering College. She earned her Ph.D. from Sathyabama University and holds an M.E. in Computer and Communication and a B.E. in Electronics and Communication Engineering from Periyar Maniammai College of Engineering, Anna University. With 17 years of teaching experience, she has published numerous papers in reputable international and national journals and at conferences. She is a life member of ISTE and a member of IEEE. Her primary areas of interest include Networks and Communication, where she actively contributes to academic and research advances.