https://jsras.rcc.edu.ly/
Vol.3 No.1 (2026), 80-91
Article history:
Received: 1 Apr. 2026
Accepted: 26 Apr. 2026
Published: 20 May. 2026

Journal of Sustainable Research in Applied Sciences

A Hybrid Framework Combining DeepLabv3+ with ResNet-50 Backbone and GOA for Automated Melanoma Lesion Segmentation

Amna Ali A Mohamed, amnaalkmati@gmail.com , https://orcid.org/0000-0001-8344-6937
Computer Engineering Department, Faculty of Information Technology, Tripoli University, Tripoli, Libya.
DOI: https://doi.org/10.36602/jsras.2026.3.1.9

Index Terms

Melanoma segmentation, DeepLabv3+, Dermoscopy, Computer-aided diagnosis, Grasshopper optimization algorithm.

Abstract

Segmenting melanoma lesions accurately is harder than it might appear. Even in high-quality dermoscopic images, lesion edges can be gradual rather than sharp, color differences between lesion and healthy skin are sometimes minimal, and no two lesions look quite alike. These are not just technical inconveniences; they directly affect how reliably a computer-aided system can support a clinician's decision. In this work, we explored whether combining deep learning with swarm intelligence could address some of these difficulties. Specifically, we paired three CNN architectures U-Net, SegNet, and DeepLabv3+ with ResNet-50, with the Grasshopper Optimization Algorithm (GOA), using it not as a preprocessing step but as a refinement stage applied after the network produces its initial prediction. The idea was that GOA's color-based clustering could tighten boundaries in regions where the CNN's confidence was low. We tested this hybrid model on the ISBI 2016 and ISIC 2018 dermoscopic datasets. Not all architectures responded equally U-Net struggled noticeably, while SegNet performed reasonably well. In previous approaches, deep learning and swarm intelligence were treated as separate stages, whereas our proposed framework, the GOA, is integrated directly into the post-prediction step of DeepLabv3+-ResNet-50, which enables boundary refinement that complements rather than replaces CNN predictions. The results from DeepLabv3+ with ResNet-50 were the best, achieving Dice coefficients of 91.7% and 88.2% and Jaccard indices of 84.9% and 79.7% on the ISBI 2016 and ISIC 2018 datasets, respectively, demonstrating the effectiveness of the proposed hybrid approach for lesion segmentation.

إطار عمل هجين يجمع بين DeepLabv3+ مع ResNet-50 Backbone وخوارزمية تحسين الجراد لتجزئة ورم الميلانوما آليا

آمنة علي عبدالسلام محمد
قسم هندسة البرمجيات، كلية تقنية المعلومات، طرابلس، ليبيا amnaalkmati@gmail.com،https://orcid.org/0000-0001-8344-6937
DOI: https://doi.org/10.36602/jsras.2026.3.1.9

الكلمات المفتاحية

تجزئة ورم الميلانوما، التعلم العميق، خوارزمية تحسين الجراد، تنظير الجلد، تحليل الصور الطبية.

الملخص

تعد التجزئة الدقيقة لورم الميلانوما خطوةً اساسية في تشخيص سرطان الجلد بمساعدة الحاسوب، إذ تؤثر بشكل مباشر على موثوقية التقييم السريري وبالتالي وضع خطة العلاج. وتمثل التجزئة الاوتوماتيكية للاورام تحديًا بسبب الاختلاف في مظهرصور الاورام، وانخفاض التباين بينها وبين الجلد المحيط، وعدم انتظام حدودها، ووجود تشوهات في الصور. ولمعالجة هذه التحديات، تقترح هذه الورقة البحثية إطار عمل هجينً بين التعلّم العميق وذكاء الأسراب لتجزئة الورم الجلدي تلقائيًا. تدمج الطريقة المقترحة خوارزمية تحسين الجراد (GOA) مع الشبكات العصبية الالتفافية (CNN)، لتحسين تحديد موقع الورم تحديد حدوده. وقد تم تقييم إطار العمل باستخدام مجموعتي بيانات ISBI 2016 وISIC 2018 المعياريتين. في الدراسات السابقة، كان يطبق التعلم العميق وذكاء الأسراب كمرحلتين منفصلتين، بينما تم دمج إطار العمل المقترح خوارزمية التحسين مباشرةً في خطوة ما بعد التنبؤ لنموذج DeepLabv3+-ResNet-50، ليُتيح تحسين الحدود بما يُكمّل تنبؤات الشبكات العصبية. وقد حققت نتائج DeepLabv3+مع ResNet-50 أفضل النتائج، حيث بلغت معاملات دايس 91.7% و88.2%، ومؤشرات جاكارد 84.9% و79.7% على مجموعتي بيانات ISBI 2016 وISIC 2018 على التوالي، ليبرهن عن فعالية الطريقة المقترحة في تجزئة ورم الميلانوما آليًا. وقد تم تقييم أداء التجزئة باستخدام الحساسية، والنوعية، والدقة، بالاضافة الى معامل دايس للتشابه (DSC)، ومؤشر جاكارد (JAC). حيث أكد التحليل النوعي كذلك قدرة اطار العمل الهجين المقترح على تحديد حدود الاورام بدقة وإنشاء أقنعة تجزئة تتطابق بشكل كبير مع ما انشاءه الخبراء. هذه النتائج تشير إلى أن دمج ذكاء الأسراب مع التعلم العميق يمكن أن يحسن بشكل كبير أداء تجزئة الورم الجلدي، ويوفر أداة واعدة لتشخيصه بمساعدة الحاسوب.

1. INTRODUCTION

Melanoma is responsible for the vast majority of skin cancer deaths despite being far less common than other skin cancer types, a disparity that underscores just how important early, accurate diagnosis is(Thapar et al., 2024). When caught at an early stage, five-year survival rates exceed 98% improving patient survival outcomes(Esteva et al., 2017). Dermoscopy has become the standard non-invasive imaging method for evaluating skin lesion (Kawahara et al., 2018). Yet translating dermoscopic images into reliable diagnoses is not straightforward. Even experienced dermatologists disagree on lesion boundaries in a meaningful proportion of cases (Codella et al., 2019), and the manual delineation of lesion regions is slow enough to be impractical in high-throughput screening contexts.

Automated segmentation using convolutional neural networks has made considerable progress in recent years. Architectures such as U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), and DeepLabv3+ (Chen et al., 2018) have demonstrated strong performance on dermoscopic benchmark datasets, and in some settings approach expert-level accuracy. That said, CNN-based methods are not without weaknesses. They can struggle when training data are limited or when lesion appearance varies substantially across patients, imaging devices, and acquisition conditions. Boundary delineation in particular remains problematic; models tend to produce smooth, averaged contours that may underperform on lesions with irregular or diffuse edges. While DeepLabv3+ with a ResNet-50 backbone has demonstrated strong performance in dermoscopic segmentation, its predictions near lesion borders are still uncertain due to inherent limitations of gradient-based training. The atrous spatial pyramid pooling mechanism can successfully capture multi-scale contextual information, but does not explicitly model the color-based spatial coherence between lesion and surrounding skin pixels. This allows for post-prediction refinement by complementary optimization methods.

One way to address these limitations is to pair deep learning with metaheuristic optimization. Swarm intelligence algorithms; including Particle Swarm Optimization (PSO) (Saifullah & Drężewski, 2024) (Mohamed et al., 2024), Ant Colony Optimization (ACO) (Sarwar et al., 2024), the Grey Wolf Optimizer (GWO) (Hu et al., 2024), and the Grasshopper Optimization Algorithm (GOA); have shown value in medical image analysis by improving feature selection and parameter tuning in ways that gradient-based training alone does not always achieve. Thapar et al. (Thapar et al., 2024) demonstrated this principle in a melanoma analysis pipeline where GOA drove the segmentation stage and a CNN was used only for downstream classification. Their results were encouraging, but the architecture kept deep learning and optimization largely separate, limiting how much each component could benefit from the other. Nevertheless, in this study GOA was chosen for several reasons. First, the nymph and adult phase behavior of GOA has the nature of balancing exploration and exploitation, which is suitable for clustering in which both broad search and fine grained refinement are needed. Secondly, GOA has shown competitive performance in image segmentation tasks with fewer parameter dependencies compared to PSO and ACO. Third, its color-based clustering approach is well-suited for dermoscopic images in which the lesion and skin regions have different but sometimes subtle color differences

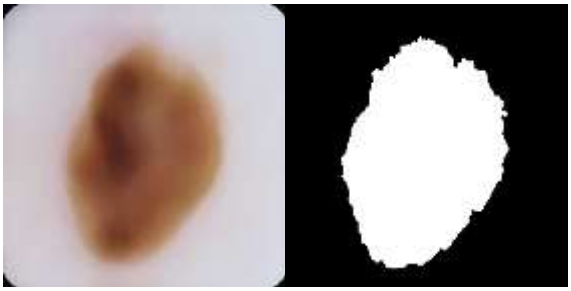
The work presented here takes a different approach. We integrate GOA directly into the output stage of a DeepLabv3+–ResNet-50 segmentation pipeline, using it to refine predicted masks rather than replace or precede the CNN. The motivation is straightforward: deep networks are good at learning discriminative features, but their predictions near lesion boundaries are often uncertain. GOA-based clustering, applied to the raw image after CNN prediction, provides complementary spatial information that can tighten those boundaries. The main contributions of this work are:

- Development of a hybrid Deep Learning–Swarm Intelligence framework for automated melanoma lesion segmentation.
- Integration of GOA with DeepLabv3+ for post-prediction boundary refinement and lesion region consistency.
- Systematic comparison of U-Net, SegNet, and DeepLabv3+ with and without GOA on two benchmark datasets.
- Evidence that combining swarm intelligence with deep segmentation networks improves robustness across varied lesion characteristics and imaging conditions.

2. MATERIAL AND METHODS

2.1 Dataset Description

This study evaluated our hybrid segmentation framework on benchmark dermoscopic datasets, including the ISBI 2016 Challenge Dataset(Codella et al., 2019) and the ISIC 2018 Dataset(Tschandl et al., 2018) . The ISBI 2016 dataset contains 900 training images and 379 test images with expert-annotated masks(Codella et al., 2019). The ISIC 2018 dataset comprises 10,015 images, covering a variety of lesion types with ground truth annotations(Tschandl et al., 2018). Images exhibit variability in size, color, and shape, reflecting real-world clinical scenarios (Esteva et al., 2017)(Kawahara et al., 2018). Figure 1 shows an image and its ground truth from ISIC 2016 Dataset images



(a) (b)

Figure 1: Dermoscopic image ISIC_ISIC_0000008 from ISIC 2016 Dataset images
(a) original image, (b) Ground Truth image

2.2 Preprocessing

Usually, melanoma skin cancer images aren't of good quality, so there's some noise in the imaging process. Therefore, those images need to be pre-processed. Moreover, dermoscopic images from the ISBI 2016 and ISIC 2018 datasets underwent preprocessing to ensure consistent input preparation for the CNN models. All images and corresponding ground truth masks were resized to a unified spatial resolution of 256×256 pixels to maintain consistent input dimensions during training and testing.

Median filtering was applied to the dermoscopic images to reduce noise and suppress unwanted artifacts while preserving important lesion boundary information. This filtering process improved image smoothness and enhanced lesion region representation.

In addition, data augmentation techniques, including horizontal reflection, were employed during training to increase dataset diversity and improve model generalization capability. These preprocessing steps contributed to more robust feature extraction and improved lesion segmentation performance

2.3 Deep Learning Module

Three convolutional neural network (CNN) architectures were implemented for initial melanoma lesion segmentation:

- **U-Net**(Ronneberger et al., 2015)**:** Fully convolutional network with symmetric encoder-decoder architecture and skip connections.
- **SegNet**(Badrinarayanan et al., 2017)**:** Encoder-decoder network using max-pooling indices to enhance segmentation accuracy.
- **DeepLabv3+ with ResNet-50 backbone**(Chen et al., 2018): Incorporates atrous spatial pyramid pooling and encoder-decoder structure for precise semantic segmentation.

All CNN models were trained using a low learning rate and mini-batch training strategy. Early stopping based on validation loss was applied to minimize overfitting and improve model generalization performance (Esteva et al., 2017) (Kawahara et al., 2018).

To further enhance lesion boundary delineation, the segmentation outputs were refined using the Grasshopper Optimization Algorithm (GOA)-based clustering framework.

The proposed framework is implemented in MATLAB using a DeepLabv3+ semantic segmentation network with ResNet-50 backbone. The dermoscopic images together with their ground-truth masks were resized to 256 × 256 pixels for processing. To improve the generalization ability of the model, data augmentation was performed by random horizontal reflections during training. The data set was randomly divided into 80% training samples and 20% validation samples. The network was trained by using training parameters including Adam optimizer with learning rate 1×10^{-3}, mini-batch size of 4, and 10 training epochs. The Grasshopper Optimization Algorithm (GOA) was used as a refining stage after the first segmentation was done by DeepLabv3+ with Resnet50. To reduce image noise before optimization, each RGB channel was passed through a median filter of size 27×27. Then, GOA-based clustering was carried out with 4 clusters (k = 4), population size of 10 grasshoppers and maximum of 7 iterations. The optimized clustering result was fused with the DeepLabv3+ with Res50 segmentation mask by a weighted fusion strategy, where the weights were 0.7 and 0.3 for CNN-based segmentation and GOA-based refinement, respectively. The resulting mask was refined by morphological operations such as hole filling and small-object removal to produce the final melanoma lesion segmentation.

2.4 Grasshopper Optimization Algorithm (GOA)

The Grasshopper Optimization Algorithm was introduced by Saremi, Mirjalili, and Lewis in 2017(Saremi et al., 2017) as a population-based optimizer modelling the collective movement of grasshopper swarms. What makes grasshopper behaviour interesting from an optimization standpoint is that it naturally alternates between two modes: slow, clustered movement during the nymph phase (which resembles exploitation of a known promising region) and rapid, dispersed flight in the adult phase (which enables broad exploration). GOA formalizes this behaviour mathematically so that the balance between the two modes can be tuned over the course of an optimization run.

Each grasshopper's position is updated according to three contributing forces; mutual social interactions between individuals, gravitational pull, and wind-driven advection:

$$X_i = S_i + G_i + A_i \tag{1}$$

where X_i is the position of the i-th grasshopper, S_i captures social interaction, Gi gravitational attraction, and A_i wind advection. The social interaction term aggregates pairwise influence across the swarm:

$$Si = \sum_{\substack{i=1 \\ i\neq j}}^{N} S(d_{ij})\ \widehat{d_{ij}} \tag{2}$$

where dij is the distance between grasshoppers i and j, and $\widehat{d_{ij}}$ is the unit vector pointing from j to i. The function S(r) governs whether two grasshoppers attract or repel each other at a given distance:

$$S(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{3}$$

Here f controls the strength of attraction and *l* sets the scale over which it acts. At short distances *S* produces repulsion, keeping individuals from collapsing to a single point; at intermediate distances it produces attraction, drawing the swarm toward promising regions; beyond a threshold distance interactions become negligible. This three-zone structure; repulsion, attraction, and comfort is illustrated in Figure 2.

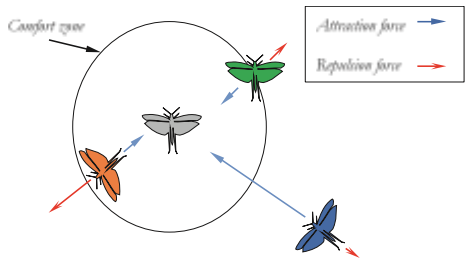


Figure 2. The forces of attraction and repulsion acting on a grasshopper

Each GOA search agent has one position vector. Where GOA update the search agent 's location based on the current search agent 's place, which is the best and also updates the position of other search agents. In equation 4, uses the factor (*C*) twice: the first time *C* from the left indicates the weight of inertia. It decreases Grasshoppers' movements toward the goal. In the second stage, it reduces areas (rest, attraction, and repulsion) between grasshoppers. Reduce the distance to locusts in their exploration and exploitation, and to be in a linear manner, as in the component $c\frac{ub_d-lb_d}{2}$ in equation(Mohamed et al., 2023). The grasshoppers are expelled from the exploration or attraction (exploitation) position to the target position using the component: $s(|x_j - x_i|)$.

$$X_i^d = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\frac{ub_d-lb_d}{2} s\left(\left|X_j^d - X_i^d\right|\right)\frac{X_j-X_i}{d_{ij}}\right) + \widehat{T_d} \quad (4)$$

The enhancements to the components enabled by decoder lesion boundary localization, which combines high-level contextual features with low-level spatial information from earlier layers, improve the quality of the extracted lesion regions and preserve structural details.

To further refine lesion regions, the Grasshopper Optimization Algorithm (GOA) is incorporated as an optimization stage following the CNN prediction process. GOA is a swarm intelligence technique based on the natural movement behavior of grasshoppers. The algorithm applies clustering-based optimization using RGB color features to separate lesion areas from surrounding healthy skin.

The optimized clustered output is merged with the CNN-generated lesion mask to produce a refined segmentation. Finally, morphological post-processing operations, including hole filling and region filtering, are subsequently applied to remove noise and improve boundary continuity. The proposed method has improved lesion localization, refined boundaries, and reduced false segmentation in

To prevent premature convergence, the algorithm progressively reduces the influence of social interactions using a shrinking coefficient c has been shown in the equation 5:

$$c = cmax - l\frac{cmax-cmin}{L} \quad (5)$$

where cmax and cmin bound the coefficient range and *l/L* is the fraction of the budget consumed where and *l* is the current iteration and *L* is the maximum number of iterations. As iterations proceed, c decreases, narrowing the effective search radius and focusing the swarm on refining a solution rather than exploring new territory(Saremi et al., 2017).

In the proposed framework, GOA is applied after initial CNN segmentation. Rather than operating on raw pixel intensities alone, it clusters RGB color features extracted from the median-filtered image to separate lesion pixels from surrounding skin. The resulting cluster labels are combined with the CNN mask through a logical OR operation, after which morphological post-processing removes isolated noise regions and fills internal holes. This design means that GOA and the CNN contribute complementary information; the network supplies learned feature-based predictions while GOA contributes color-driven spatial coherence rather than one simply replacing the other.

2.5 Hybrid DeepLearning GOA Integration

Our proposed method for the melanoma lesion analysis framework combines DeepLabv3+ with a ResNet-50 backbone and the Grasshopper Optimization Algorithm (GOA) to to extract skin lesions from dermoscopic images. DeepLabv3+ employs an encoder–decoder architecture in which ResNet-50 serves as the feature-extraction network(Sriwiboon, 2025). The residual connections in ResNet-50 support deep feature learning while reducing degradation problems during training. The framework also incorporates Atrous Spatial Pyramid Pooling (ASPP) to capture multiscale contextual information through multiple atrous convolution rates, enabling detection of melanoma lesions with varying shapes, textures, and sizes.

dermoscopic images from the ISIC 2016 and ISIC 2018 datasets.

2.6 Workflow of the Proposed Method

The workflow of the proposed melanoma lesion analysis framework comprises several sequential stages that integrate deep learning models with the Grasshopper Optimization Algorithm (GOA) for lesion refinement. Our proposed method is illustrated in Figure 3.

The dermoscopic images from the ISBI 2016 and ISIC 2018 datasets were collected and resized to a unified input resolution to ensure consistent processing. The ground-truth mask for each image was included for supervised training and evaluation.

The three convolutional neural network architectures, namely U-Net, SegNet, and DeepLabv3+ with ResNet-50 backbone, were trained independently in the first stage,to generate initial lesion masks. during the training stage, data augmentation techniques were applied, which improved model robustness and reduced overfitting. Each

CNN model learned discriminative lesion features from dermoscopic images and produced preliminary lesion segmentation outputs.The post-processing refinement stage based on the Grasshopper Optimization Algorithm (GOA) was applied after obtaining the initial segmentation masks. During optimization, median filtering was utilized to reduce noise and smooth the dermoscopic image while preserving important lesion structures. RGB color features were then extracted from the filtered images and used as inputs for GOA-based clustering. In this stage, the GOA process searched for optimal cluster centers that could distinguish lesion regions from surrounding healthy skin. The clustering output generated was combined with the CNN-predicted lesion masks to refine lesion boundaries and improve region consistency.

Several procedures were applied to improve segmentation quality, including morphological operations, hole filling, image closing, and removal of small isolated regions. These operations reduced noise artifacts and improved lesion continuity. At the end, the evaluations were performed by comparing the refined lesion masks with their ground-truth masks. The Sensitivity, Specificity, Accuracy, Dice coefficient, and Jaccard index. The framework was evaluated on both ISBI 2016 and ISIC 2018 datasets to assess segmentation robustness and generalization capability across different lesion appearances and imaging conditions.

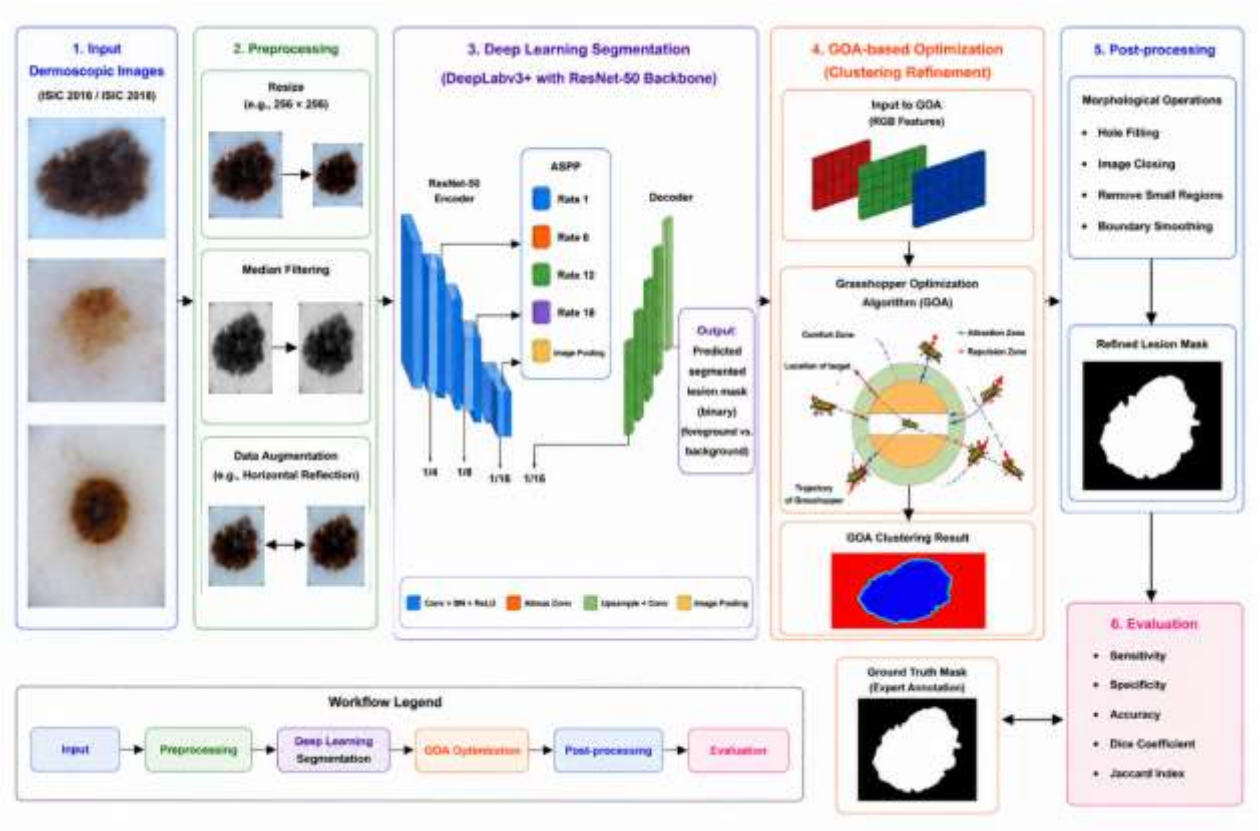


Figure 3. The proposed method combining DeepLabv3+ with ResNet-50 + GOA for Melanoma Lesion Analysis

3. Experimental Results

3.1 Evaluation Metrics

The proposed method's performance was measured using a confusion matrix, which uses five standard metrics derived from pixel-level confusion matrix counts: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), where positive refers to lesion pixels and negative to healthy skin pixels (Thapar et al., 2024) (Codella et al., 2019). The relationship between these quantities and the evaluation metrics is illustrated in Figure 4.

The following evaluation metrics were used in the experiment.

- **Accuracy(ACC):** measures the proportion of all pixels ,both lesion and background , that were correctly classified. It is a useful overall indicator, but can be misleading when lesion pixels represent a small fraction of the image, since a model that labels everything as background would still score well. As illustrated in Figure 4(a) accuracy provides an overall measure of segmentation performance. **Sensitivity(SEN):** captures how completely the model identifies the lesion region. Low sensitivity means the model misses parts of the true lesion. This metric is also called recall or The true positive rate(Müller et al., 2022).
- **Specificity(SPE):** This quantitative metric measures the model's ability to exclude healthy skin from the predicted mask. High specificity with low sensitivity is a common failure mode in segmentation, where the model is overly conservative. Sensitivity and specificity metrics are visualized in the Figure 4(b) and Figure 4(c).
- **The Dice Similarity Coefficient (DSC):** DSC is arguably the most informative single metric for segmentation tasks. It measures the overlap between predicted and ground-truth masks as twice the intersection divided by the sum of their sizes, penalising both missed lesion pixels and spurious predictions equally. DSC ranges from 0 (no overlap) to 1 (perfect match). Figure 4(d) illustrates the Dice coefficient.
- **The Jaccard Index(JAC):** JAC is also known as Intersection over Union (IoU). It is related to DSC but applies a stricter penalty, dividing the intersection by the union rather than the average size. As illustrated in Figure 4(e), for any given prediction, Jaccard will always be lower than or equal to the corresponding Dice score, making it a more conservative measure of segmentation quality (Jabber et al., 2025).

The performances of the proposed framework are quantitatively evaluated with standard segmentation metrics such as Accuracy, Sensitivity, Specificity, Dice Coefficient and Jaccard Index. These metrics are calculated by comparing the predicted segmentation masks to the ground-truth annotations available in the dataset. The performance measurements formulations are summarised in Table 1.

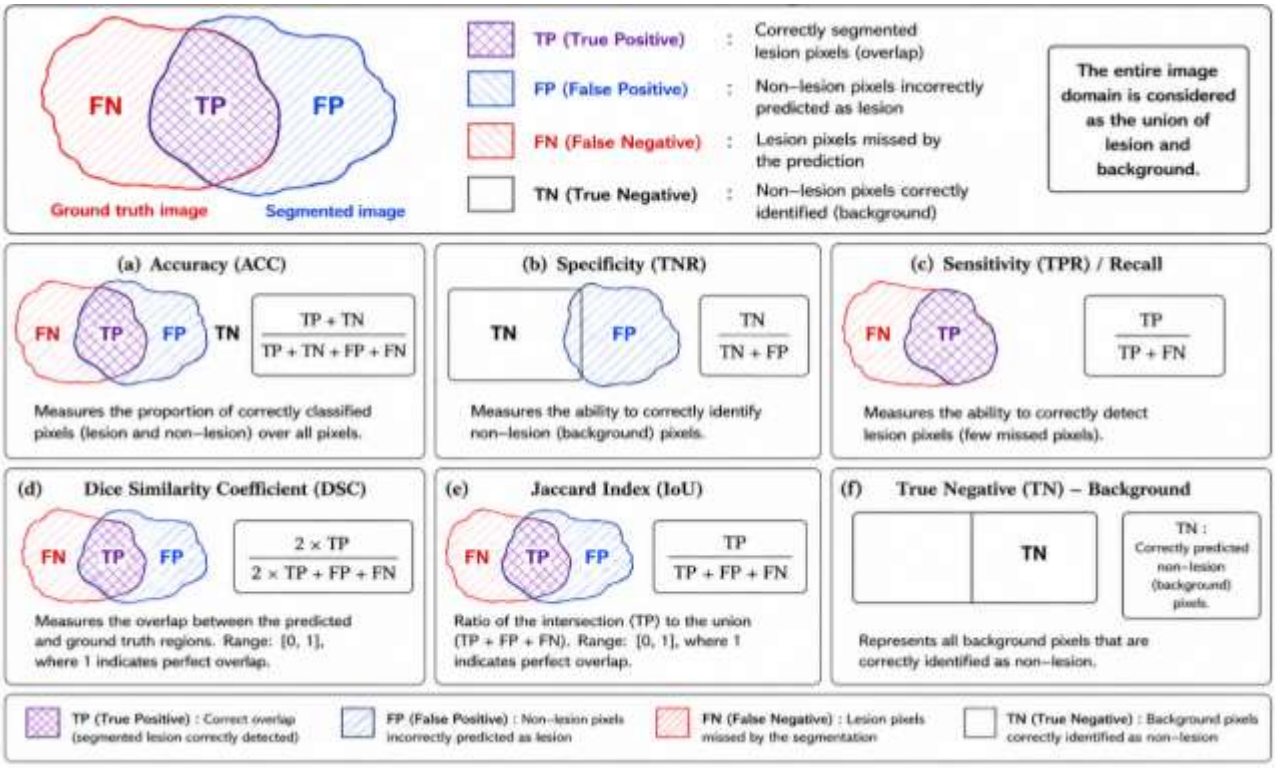


Figure 4. Visual interpretation of the evaluation metrics used for melanoma lesion segmentation.
(a) Accuracy, which considers correctly classified lesion and background pixels (TP and TN); (b) Specificity, which measures correct identification of background pixels (TN); (c) Sensitivity, which measures correct identification of lesion pixels (TP); (d) Dice Similarity Coefficient (DSC), illustrating the overlap between the segmented lesion and the ground-truth lesion; and (e) Jaccard Index (JAC), representing the ratio between the intersection and union of the segmented and ground-truth regions

Table 1: The performance metric

Measure	Formula
Accuracy	(TP + TN) / (TP + FP+ TN + FN)
Specificity	TN / (FP + TN)
Sensitivity	TP / (TP + FN)
Dice coefficient	2*TP / (2*TP + FP + FN)
Jaccard index	TP / (TP + FP + FN)

3.2 Quantitative Results

- ***Segmentation accuracy analysis***

We have evaluated three different CNN architectures: U-Net, SegNet, and DeepLabv3+ResNet50 to investigate the impact of the Grasshopper Optimization Algorithm (GOA) on lesion segmentation performance. Tables 2 and 3 summarize the quantitative results obtained on the ISBI 2016 and ISIC 2018 datasets, respectively. The performance has been evaluated on the ISBI 2016 and ISIC 2018 test sets for all CNN architectures (U-Net, SegNet, DeepLabv3+ with ResNet-50) using the swarm intelligence optimization algorithm GOA. The evaluation was performed using the Dice coefficient, Jaccard index, sensitivity, specificity, and accuracy.

Table 2: Results of our proposed method using different CNN models on the ISBI 2016 dataset

Model	Sen(%)	Spe(%)	Acc(%)	Dsc(%)	Jac(%)
U-Net	38.4	99.3	80.4	47.2	36.4
Seg-Net	87.1	96.3	92.6	87.6	78.9
DLabv3+ ResNet50	92.7	97.2	97.9	91.7	84.9

Table 3: Results of our proposed method using different CNN models on the ISIC 2018 dataset

Model	Sen(%)	Spe(%)	Acc(%)	Dsc(%)	Jac(%)
U-Net	54.5	96.1	87.2	57.6	47.8
Seg-Net	69.5	98.5	89.7	72.9	64.1
DLabv3+ ResNet50	82.8	99.6	97.0	88.2	79.7

Table 2 shows that the performance of the three CNN architectures varied considerably. U-Net achieved the lowest performance, obtaining a Dice coefficient of 47.2% and a Jaccard index of 36.4%, indicating limited overlap between the predicted segmentation and the ground-truth mask. Using SegNet has improved the segmentation performance. The Dice coefficient was 87.6%, while the other was 78.9%. This improvement indicates that the SegNet achieves better results for specifying lesion boundaries. Among all evaluated architectures, DeepLabv3+ with ResNet-50 achieved the best overall performance. The model achieved a sensitivity of 92.7%, specificity of 97.2%, accuracy of 97.9%, Dice coefficient of 91.7%, and Jaccard of 84.9%. The findings show a high degree of agreement between the predicted masks and the ground truth, and the effectiveness of combining DeepLabv3+ with GOA for melanoma lesion segmentation. The other results were applied by using the ISIC 2018 dataset. They yielded more challenging results as presented in Table 3. U-Net achieved the lowest segmentation performance, with a Dice coefficient of 57.6% and a Jaccard index of 47.8%, whereas SegNet produced improved results, reaching a Dice coefficient of 72.9% and a Jaccard index of 64.1%, confirming its better segmentation ability compared with U-Net.

On the ISIC 2018 dataset, DeepLabv3+ with ResNet-50 again gave the strongest results: sensitivity 82.8%, specificity 99.6%, accuracy 97.0%, Dice 88.2%, Jaccard 79.7%. The specificity is worth pausing on, 99.6% means the model was almost never fooled into labeling normal skin as a lesion. The Dice and Jaccard scores are solid, though not perfect, and that gap is honest because ISIC 2018 is a genuinely harder dataset.

3.3 Qualitative Results

Beyond the numbers, we wanted to see what the segmentation actually looked. Metrics like Dice and Jaccard indicate overlap, but they do not always capture whether a mask looks clinically reasonable.

Figures 5 and 6 show representative examples from the ISBI 2016 and ISIC 2018 datasets, respectively. Each row displays, left to right: the original dermoscopic image, the GOA clustering output, the predicted mask overlaid on the image, the final binary mask, and the corresponding expert annotation.

A few things stand out from these visual results. First, the GOA clustering stage does genuine work, it extracts the lesion-dominant color region, yielding a final mask with better coverage than the CNN prediction alone, particularly around the lesion periphery, where network confidence tends to drop. Second, the overlay images show that the predicted contours track the expert annotations reasonably closely across a range of lesion types, if small or large, roughly circular and highly irregular, high-contrast and low-contrast alike.

Not every case was clean. Lesions with diffuse edges or dense hair coverage were the hardest for the framework boundary predictions in these images were less precise, and GOA could not fully recover them.

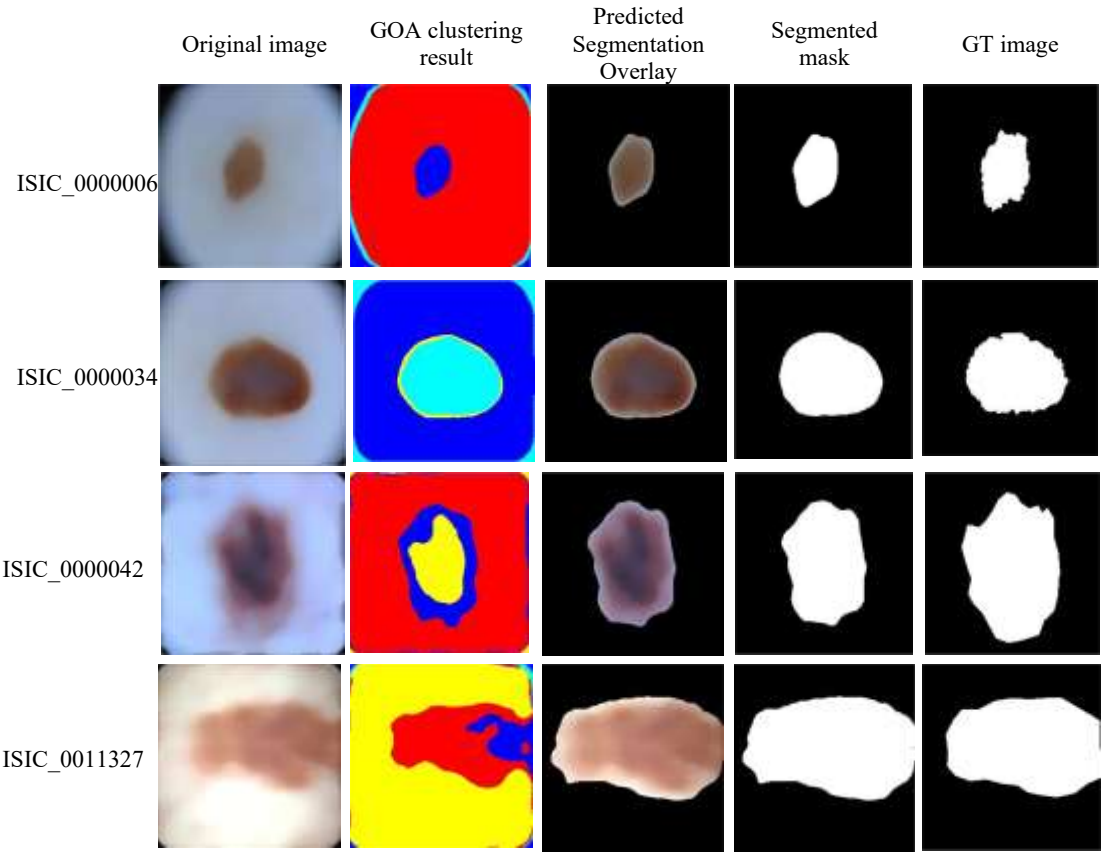


Figure 5. Visual comparison of melanoma lesion segmentation results from ISBI 2016 dataset

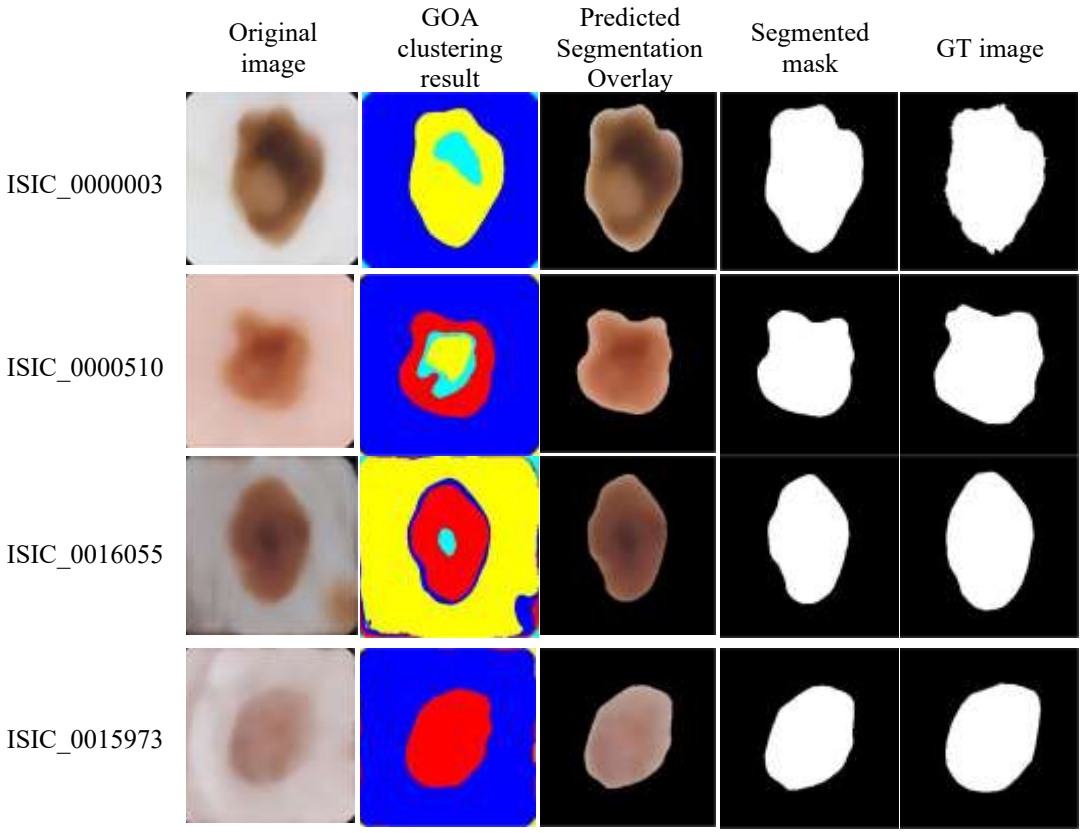


Figure 6. Visual comparison of melanoma lesion segmentation results from ISIC 2018 dataset

These cases appear in the figures without filtering, because selective presentation would paint an unrealistically positive picture. For most lesions, though, the predicted masks tracked the expert annotations well enough to be practically useful.

3.4 Comparative Analysis

In this study, we have implemented various CNN models to compare with our proposed hybrid model; these models are shown in the Tables 4 and 5. The tables compare U-Net, SegNet, and DeepLabv3+ with ResNet-50 without GOA refinement, giving a clearer picture of what each architecture contributes on its own before optimization is applied.

Table 4: Performance evaluation of CNN models on the ISBI 2016 dataset

Model	Sen(%)	Spe(%)	Acc(%)	Dsc(%)	Jac(%)
U-Net	32.1	99.5	79.0	41.9	30.9
Seg-Net	85.9	96.3	91.5	84.5	75.6
DLabv3+ ResNet50	89.6	97.3	93.9	89.9	82.8

Table 5: Performance evaluation of CNN models on the ISIC 2018 dataset

Model	Sen(%)	Spe(%)	Acc(%)	Dsc(%)	Jac(%)
U-Net	46.9	96.5	85.5	51.9	41.1
Seg-Net	66.7	98.6	89.6	70.1	61.6
DLabv3+ ResNet50	74.0	99.5	90.7	80.8	70.9

The gap between U-Net and the other two architectures is striking, and worth discussing rather than simply reporting. At ISBI 2016, U-Net achieved only a 32.1% sensitivity despite a specificity of 99.5% a pattern that tells a specific story. The model was extremely reluctant to label anything as a lesion, which kept its false positive rate near zero but caused it to miss large portions of actual lesion tissue. This kind of over-conservative behavior is not unusual when a symmetric encoder-decoder is trained on a relatively small, class-imbalanced dataset, and it suggests that the U-Net in this configuration was not well calibrated for the task.

SegNet recovered much of that lost sensitivity, reaching 85.9% on ISBI 2016, and its Dice score of 84.5% represents a substantial jump from U-Net's 41.9%. The max-pooling index mechanism, which distinguishes SegNet from U-Net, appears to help it recover finer spatial detail during decoding, which is important for lesion boundary reconstruction.

DeepLabv3+ with ResNet-50 pushed performance further still. On ISBI 2016, it reached a Dice of 89.9% and a Jaccard of 82.8%, and on the harder ISIC 2018 dataset, it achieved 80.8% and 70.9% consistently, the best figures across both datasets and all metrics. The atrous convolutions in DeepLabv3+ allow the model to aggregate context across multiple scales without losing spatial resolution, which seems particularly well-suited to dermoscopic segmentation where lesion size varies considerably across images. The ResNet-50 backbone also provides deeper residual feature learning than the shallower encoders in U-Net and SegNet, which likely contributes to the performance gap on the more varied ISIC 2018 cases.

As shown in the tables, specificity is higher than sensitivity across all results from both datasets' models. That means the substantial class imbalance between lesion and background pixels in dermoscopic images, which influences predictions, pushing models toward caution. Addressing this imbalance more aggressively during training, through weighted loss functions or oversampling strategies, may be a productive avenue for future work.

To further evaluate the proposed framework against existing approaches in the literature, Tables 6 and 7 present a comparative analysis

Table 6: Comparison with State-of-the-Art Segmentation Methods on the ISBI 2016 Dataset

Reference	Year	Method	SEN(%)	SPE (%)	ACC(%)	DSC (%)	JAC(%)
(Gutman et al., 2016)	2016	Team-EXB	91.0	96.5	95.30	91.00	84.30
(Garcia-Arroyo & Garcia-Zapirain, 2017)	2017	Fuzzy Classification	87.0	-	93.4	86.9	79.1
(Jahanifar et al., 2019)	2019	Saliency Map	90.1	98.2	94.3	90.7	83.8
(Tang et al., 2019)	2019	MS-UNet	91.36	95.98	95.3	89.91	83.22
(Wang et al., 2022)	2020	CCA-Net	-	-	96.1	92.6	87.1
(Bukhari et al., 2023)	2023	Depthwise-Net	92.0	90.0	95.0	-	-
Proposed	**2026**	**DeepLabv3+ ResNet50+GOA**	**92.7**	**97.2**	**97.9**	**91.7**	**84.9**

Table 7: Comparison with State-of-the-Art Segmentation Methods on the ISIC 2018 Dataset

Reference	Year	Method	SEN(%)	SPE(%)	ACC(%)	DSC (%)	JAC(%)
(Shahin et al., 2019)	2019	PPM Encoder-Decoder	90.2	97.4	-	90.30	83.70
(Alhudhaif et al., 2022)	2022	Fusion Loss U-Net	94.0	94.0	93.0	88.00	80.00
(Innani et al., 2023)	2023	GAN-based MGAN	93.6	95.5	94.5	90.1	83.6
(Jabber et al., 2025)	2025	SkinAttn-Net	83.0	93.0	91.43	-	81.0
(Li et al., 2025)	2025	HyperFusion-Net	92.4	-	-	93.0	88.1
Proposed	**2026**	**DeepLabv3+ ResNet50+GOA**	**82.8**	**99.6**	**97.0**	**88.2**	**79.7**

These tables present a comparison of the proposed framework with representative state-of-the-art segmentation methods evaluated on the ISBI 2016 and ISIC 2018 datasets is shown. Only methods performing lesion segmentation are included for a fair and direct comparison. The proposed DeepLabv3+-ResNet50+GOA framework achieves competitive performance on both datasets. The proposed method achieves the highest accuracy of 97.9% and competitive Dice and Jaccard scores compared to all the compared methods at ISBI 2016. The proposed method achieves a high accuracy of 97.0% on ISIC 2018. Some recent architectures with attention mechanisms and transformer-based components have reported higher Dice and Jaccard scores (as discussed further in the Discussion section).

Table 7 shows that the proposed framework achieves the best specificity of 99.6% and a competitive accuracy of 97.0%. However, recent attention-based architectures such as HyperFusion-Net(Li et al., 2025) have shown better Dice and Jaccard scores. This is discussed further in the Discussion.

4. DISCUSSION AND CONCLUSION

The results of this study demonstrate that combining deep learning and swarm intelligence yields measurable improvements in automated melanoma lesion segmentation. Among the three architectures evaluated, DeepLabv3+ with ResNet-50 was by far the strongest and its combination with GOA produced additional gains, consistently the most at lesion boundaries where CNN predictions are inherently uncertain.

From a clinical perspective, the Dice coefficient of 91.7% achieved on the ISBI 2016 dataset is clinically meaningful. In dermoscopic practice, accurate lesion boundary delineation directly influences treatment planning: an undersegmented mask may cause a surgeon to underestimate the excision margin required, while an oversegmented mask may lead to unnecessarily wide excisions. A Dice score above 90% indicates that the predicted mask overlaps with the expert annotation to a degree that approaches inter-observer variability between experienced dermatologists, which has been reported in the range of 85%-92% in the literature (Gutman et al., 2016). This suggests that the proposed framework produces clinically usable segmentations on well-standardized dermoscopic data.

On the ISIC 2018 dataset, the Dice of 88.2% and Jaccard of 79.7% reflect the genuine difficulty of that benchmark, which contains more lesion type diversity, more imaging variability, and more ambiguous boundary cases. The high specificity of 99.6% on ISIC 2018 is noteworthy: it means the model almost never labels healthy skin as a lesion, a desirable property in a computer-aided diagnosis tool because false positive predictions tend to increase unnecessary clinical follow-up. The lower sensitivity (82.8%) indicates that some lesion pixels were missed, and addressing this trade-off through improved class balancing during training is identified as a direction for future work.

The qualitative findings confirm the quantitative results. GOA-based clustering provides genuine boundary improvement, recovering lesion periphery coverage where CNN predictions were conservative. Challenging cases involving diffuse boundaries or heavy hair artifact, which represent known failure modes of CNN-based dermoscopic segmentation, were retained in the reported figures without filtering.

From the comparison with state-of-the-art methods in Tables 6 and 7, the proposed framework achieves the highest accuracy in both datasets among all compared methods. Recent architectures with attention mechanisms and transformer-based components, such as HyperFusion-Net (Li et al., 2025) have reported higher Dice and Jaccard scores on ISIC 2018. This gap is honest and expected: transformer-based models have more parameters, benefit from pre-training on large scale vision datasets, and are architecturally better suited to detect long-range spatial dependencies. The contribution of this work is not to outperform all transformer-based architectures but to show that a well-integrated CNN-swarm intelligence hybrid can achieve a competitive accuracy with a considerably simpler and more interpretable architecture.

The U-Net has a very low Dice of 47.2% on ISBI 2016 due to very high specificity and very low sensitivity. This indicates the model defaulted to predicting background due to class imbalance during training. This highlights the need for explicit class-balancing strategies regardless of the architecture used.

4.1 Limitations

Despite the promising results reported in this study, several limitations should be acknowledged.

First, formal statistical significance testing was not conducted. The performance comparisons between

architectures and between CNN-only and CNN+GOA configurations are based on single test-set evaluations rather than repeated experiments or cross-validation folds. Without statistical tests such as Wilcoxon signed-rank tests or paired t-tests, it is not possible to confirm that the observed performance differences are statistically significant rather than due to variability in the test partition. It should be noted, however, that statistical significance testing of this kind is most applicable in experimental designs where multiple independent runs or cross-validation folds produce a distribution of performance scores that can be compared. In the present study, the evaluation follows the standard protocol established by the ISBI 2016 and ISIC 2018 benchmark challenges, in which all competing methods are evaluated on fixed, predefined test sets with a single inference pass. This protocol is universally adopted in the dermoscopic segmentation literature including all state-of-the-art methods included in Tables 6 and 7 and does not produce the repeated measurements required as input to Wilcoxon signed-rank or paired t-tests. Applying such tests to single test-set scores would therefore be statistically inappropriate rather than merely omitted.

Second, the GOA parameters, four clusters, population size of ten, and a maximum of seven iterations were experimentally determined instead of through systematic ablation, meaning that alternative settings could lead to better performance.

Third, the evaluation was conducted only on the ISBI 2016 and ISIC 2018 datasets that may not sufficiently cover the imaging variability across different hospitals, dermatoscope models, and patient populations, limiting generalisability to out-of-distribution data.

Fourth, class imbalance between lesion and background pixels was not explicitly addressed during training beyond standard data augmentation, which contributed to the over-conservative predictions observed in U-Net.

Finally, the implementation was developed in MATLAB, which can be limiting in terms of accessibility and reproducibility for researchers using other frameworks such as Pytorch. The source code was not publicly released.

4.2 Recommendations and Future Work

Based on the findings of this study, the following recommendations are proposed for researchers and practitioners working in automated melanoma segmentation:

First, the GOA parameter configuration used in this paper (k=4 clusters, population size of 10, 7 iterations) was determined based on empirical testing. A systematic ablation study over these parameters would help to establish more principled configuration guidelines and better understand the sensitivity of the framework to hyperparameter choices.

Second, future work should explore the use of transformer-based backbones, e.g., Swin Transformer or Vision Transformer (ViT), to replace ResNet-50 in the DeepLabv3+ framework. The performance gap observed on ISIC 2018 compared to attention-based architectures indicates that incorporating long-range spatial context modeling could be a promising approach to improve segmentation on more heterogeneous datasets.

Third, validation on additional and more recent dermoscopic datasets, including multi-center and multi-device collections would enhance the generalisability claims of the framework. The current evaluation is restricted to ISBI 2016 and ISIC 2018, which are established benchmarks, but do not fully capture the imaging variability in real clinical deployment across different hospital settings and device types.

Finally, clinical validation studies with dermatologists reviewing the outputs of the framework in conjunction with their own annotations would provide practical insight into whether the proposed system meets a threshold of clinical utility, and would help identify failure cases that are most consequential from a patient safety perspective.

In summary, the scientific contributions of this study are as follows: (1) a hybrid framework is proposed and validated, in which GOA is directly integrated into the post-prediction stage of DeepLabv3+–ResNet-50 instead of treating deep learning and swarm intelligence as sequential independent steps; (2) a systematic comparison of three CNN architectures with and without GOA is provided on two benchmark datasets to quantify the contribution of each component; and (3) it is shown that GOA-based color clustering is a computationally efficient complement to gradient-trained segmentation networks, particularly for boundary refinement of dermoscopic images in which CNN predictions are most uncertain.

REFERENCES

Alhudhaif, A., Ocal, H., Barisci, N., Atacak, İ., Nour, M., & Polat, K. (2022). A novel approach to skin lesion segmentation: Multipath fusion model with fusion loss. *Computational and Mathematical Methods in Medicine*, *2022*(1), 2157322.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495.

Bukhari, M., Yasmin, S., Habib, A., Cheng, X., Ullah, F., Yoo, J., & Lee, D. (2023). A novel framework for melanoma lesion segmentation using multiparallel depthwise separable and dilated convolutions with swish activations. *Journal of Healthcare Engineering*, *2023*(1), 1847115.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., & Marchetti, M. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *ArXiv Preprint ArXiv:1902.03368*.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S.

M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.

Garcia-Arroyo, J. L., & Garcia-Zapirain, B. (2017). Segmentation of skin lesions based on fuzzy classification of pixels and histogram thresholding. *ArXiv Preprint ArXiv:1703.03888*.

Gutman, D., Codella, N. C. F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *ArXiv Preprint ArXiv:1605.01397*.

Hu, P., Han, Y., & Zhang, Z. (2024). Multi-level thresholding color image segmentation using modified gray wolf optimizer. *Biomimetics*, *9*(11), 700.

Innani, S., Dutande, P., Baid, U., Pokuri, V., Bakas, S., Talbar, S., Baheti, B., & Guntuku, S. C. (2023). Generative adversarial networks based skin lesion segmentation. *Scientific Reports*, *13*(1), 13467.

Jabber, A. A., Shadeed, G. A., Salim, N., & Dibs, H. (2025). Automated skin lesion diagnosis and classification using k-mean, lab-color-space segmentation and deep learning. *Operations Research Forum*, *6*(2), 43.

Jahanifar, M., Tajeddin, N. Z., Asl, B. M., & Gooya, A. (2019). Supervised Saliency Map Driven Segmentation of Lesions in Dermoscopic Images. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 509–518. https://doi.org/10.1109/JBHI.2018.2839647

Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2018). Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 538–546.

Li, M., Jiang, Y., Cao, G., Xu, T., & Guo, R. (2025). HyperFusionNet combines vision transformer for early melanoma detection and precise lesion segmentation. *Scientific Reports*.

Mohamed, A. A. A., Hançerlioğullari, A., Rahebi, J., Ray, M. K., & Roy, S. (2023). Colon Disease Diagnosis with Convolutional Neural Network and Grasshopper Optimization Algorithm. In *Diagnostics* (Vol. 13, Issue 10). https://doi.org/10.3390/diagnostics13101728

Mohamed, A. A. A., Rahebi, M., Hançerlioğulları, A., & Rahebi, J. (2024). An approach based on convolutional neural network and ACO-PSO for colon cancer disease diagnosis. *Politeknik Dergisi*, *28*(2), 649–659.

Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net:Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Saifullah, S., & Drężewski, R. (2024). Advanced medical image segmentation enhancement: A particle-swarm-optimization-based histogram equalization approach. *Applied Sciences*, *14*(2), 923.

Saremi, S., Mirjalili, S., & Lewis, A. (2017). Grasshopper optimisation algorithm: theory and application. *Advances in Engineering Software*, *105*, 30–47.

Sarwar, N., Irshad, A., Naith, Q. H., D. Alsufiani, K., & Almalki, F. A. (2024). Skin lesion segmentation using deep learning algorithm with ant colony optimization. *BMC Medical Informatics and Decision Making*, *24*(1), 265.

Shahin, A. H., Amer, K., & Elattar, M. A. (2019). Deep convolutional encoder-decoders with aggregated multi-resolution skip connections for skin lesion segmentation. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 451–454.

Sriwiboon, N. (2025). A high-performance deep learning framework for skin lesion segmentation based on DeepLabV3+ and multi-scale context modeling. *Biomedical Materials & Devices*, 1–17. https://doi.org/10.1007/s44174-025-00497-w

Tang, Y., Yang, F., Yuan, S., & Zhan, C. (2019). A multi-stage framework with context information fusion structure for skin lesion segmentation. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 1407–1410.

Thapar, P., Rakhra, M., Alsaadi, M., Quraishi, A., Deka, A., & Ramesh, J. V. N. (2024). A hybrid Grasshopper optimization algorithm for skin lesion segmentation and melanoma classification using deep learning. *Healthcare Analytics*, *5*, 100326.

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, *5*(1), 180161.

Wang, R., Chen, S., Ji, C., & Li, Y. (2022). Cascaded context enhancement network for automatic skin lesion segmentation. *Expert Systems with Applications*, *201*, 117069.