Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 7 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i07.54621

PAPER

DeepSkinNet: A Deep Learning Induced Skin Lesion Extraction System from Dermoscopic Images

Abhipsa Pattanaik(✉), Leena Das

School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India

2181057@kiit.ac.in

ABSTRACT

In this work, a DeepSkinNet model was developed based on an encoder-decoder type framework. The designed encoder incorporates three blocks where each block sandwiches convolution, rectified linear unit (ReLU), and maxpooling layers to retain the prominent details. The developed DIL (dilated convolution + instance normalization + Leaky ReLU) module comprises three branches, where each branch consists of an atrous convolution layer with a sampling rate of two, followed by instance normalization and a Leaky ReLU activation function to retain the subtle details accurately. Further, the proposed decoder network with a feature fusion mechanism stacks convolution, transposed convolution, and ReLU activation functions to precisely extract the lesion regions from the dermoscopic images. The efficacy of the designed DeepSkinNet is validated through subjective as well as objective analysis and found to be suitable for medical diagnosis against various SOTA methods. The Dice coefficients (DC) found using ISIC 2016, ISIC 2017, ISIC 2018, and PH^2 datasets are found to be 93.33%, 89.00%, 92.05%, and 91.24%, respectively.

KEYWORDS

Skin lesion, segmentation, DNN, feature fusion

1 INTRODUCTION

Skin lesion extraction is a crucial step in automated dermatological analysis. It aids in isolating the lesion from surrounding healthy skin, reducing noise, and improving the accuracy of subsequent classification or segmentation models. By accurately extracting the lesion, we enhance feature extraction and ensure more reliable diagnosis, particularly in computer-aided detection systems for melanoma and other skin diseases such as skin cancer. Skin cancer is one of the major health problems in the world today, and its early detection and diagnosis are vital for proper treatment. Dermoscopy is an increasingly popular, non-invasive technique that has emerged as a significant tool for the dermatologist in magnifying the skin surface for detailed images [1]. Manual analysis or visual inspection of dermoscopic images is

Pattanaik, A., Das, L. (2025). DeepSkinNet: A Deep Learning Induced Skin Lesion Extraction System from Dermoscopic Images. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(7), pp. 15–28. https://doi.org/10.3991/ijoe.v21i07.54621

Article submitted 2025-01-27. Revision uploaded 2025-04-01. Final acceptance 2025-04-01.

an onerous and subjective process and may be subject to human error. To overcome such limitations, computer-aided diagnostics (CAD) systems have been developed to aid dermatologists in the accurate identification and classification of skin lesions [2]. Moreover, these systems make the complete diagnosis system faster than the manual examinations via dermatologist, making it one of the essential diagnostic tools for co-assisting the experts.

One of the key elements of these CAD systems is the correct extraction of the skin lesions from the surrounding healthy skin for effective therapeutics. However, this task is challenging due to various factors present within the dermoscopic images, such as illumination variations, hair follicles, air bubbles, and irregular lesion shapes. The existing methods have difficulty in effectively handling these complexities, which causes inaccurate lesion extraction results [3] [4].

For overcoming these challenges, a unique model called DeepSkinNet was introduced for robust and accurate extraction of skin lesions. Novelties within DeepSkinNet architecture include the following:

- A novel encoder network is proposed comprising a convolution layer, ReLU activation function, and max-pooling layer. The convolutional layer can maintain the spatial relation among the neighborhood pixels effectively. Also, the ReLU activation function introduces non-linearity by learning the negative values as zero, which can reduce overfitting. Further, the max-pooling layer learns the prominent features within the pooling area. The hybridization of the convolutional layer, ReLU layer, and max-pooling layer can extract fine-scale details from the lesion regions of dermoscopic scans effectively. The developed DIL module, consisting of a dilated convolution layer and instance normalization followed by a ReLU activation function, can retain the coarse scale subtle details efficiently.
- The proposed DIL module sandwiches the atrous convolution layers followed by the instance normalization layer and LeakyRelu activation function to extract the dense details within the dermoscopic image accurately. Also, the DIL module maintains the contextual relationship among the pixels by utilizing the receptive field. Further, the DIL module acts as a bridge between the encoder-decoder networks for data transmission while maintaining the spatial relationship.
- The proposed decoder network stacks the convolution, transposed convolution, and ReLU activation function that learns a mapping from feature level to pixel level for an effective lesion extraction mask.
- The developed DeepSkinNet architecture is trained on fewer samples of the ISIC 2016 dataset while it is tested on the ISIC 2017, ISIC 2018, and PH^2 dataset, which validate the efficacy of the designed model in an unseen setup.
- The developed DeepSkinNet has 6.9 M parameters, indicating it is a lightweight network, and it surpasses 18 recently developed SOTA approaches, which may be suitable for real-time lesion diagnosis.

This work aims to contribute towards designing more accurate and efficient CAD systems within skin cancer to realize an earlier detection and treatment process for this critical health issue.

The rest of the paper is organized as follows: Section 2 discusses the existing SOTA approaches developed so far. Section 3 provides a detailed description of the proposed technique. Section 4 provides a glimpse of the datasets used for the experimentation work. Section 5 discusses the subjective and objective evaluation of the developed framework. Finally, the conclusion of the paper is provided in Section 6.

2 LITERATURE REVIEW

Skin lesion extraction using a computer vision approach has a significant role in the effective and timely detection of melanoma, thereby improving patient care. Numerous lesion extraction techniques using computer vision are demonstrated by researchers worldwide, which may be segregated as unsupervised approaches and supervised approaches. Supervised approaches [5], [6] utilizing convolutional neural networks (CNN) have gained focus due to their unparalleled feature extraction capabilities, scalability, and accuracy [7]. Here, the literature review was limited to recent skin lesion extraction approaches.

In recent years, numerous researchers have developed various DNN-based models for skin lesion extraction [8]–[12]. A deep learning-based computer-aided system [8] detects and classifies skin diseases using the HAM-10000 dataset, achieving 98.3718% accuracy. While offering high accuracy, it may require significant computational resources for real-time use. The DSC-EDLMGWO model [9] integrates SE-DenseNet with an ensemble of LSTM, ELM, and SSDA, optimized using Gray Wolf Optimization, achieving 98.38% and 98.17% accuracy on HAM10000 and ISIC datasets. While highly accurate, it demands substantial computational resources and complex tuning. The RCS-ECNN method [10] integrates random cat swarm optimization with an ensemble CNN, using DNN, KDNN, GrabCut segmentation, and feature extraction for skin cancer classification, achieving 99.56% accuracy on HAM10000 and ISIC datasets. While highly accurate, it may require high computational resources and careful parameter tuning. The MobileNet-Optimized Attention Transfer framework [11] integrates self-attention, cross-attention, and an Optical Microscope Algorithm for hyperparameter tuning, achieving 98.89% accuracy with reduced computational time (1.53 sec) on ISIC and HAM10000 datasets. While highly efficient and scalable for clinical use, its performance may depend on dataset variability and model generalization. The study [12] employs a tailored preprocessing pipeline with EfficientNet B0-B7 models using transfer learning from ImageNet weights, enhancing skin cancer classification accuracy. EfficientNet-B7 achieves top-1 accuracy of 84.4% and top-5 accuracy of 97.1%, but challenges remain in classifying certain classes such as eczema and psoriasis. In order to efficiently capture long-range relationships and global context information for lesion extraction, FAT-Net [13] incorporates a transformer branch into the conventional encoder-decoder architecture. It may struggle with high computational complexity and real-time processing constraints. Tang et al. [14] propose a Separable-U-Net framework that integrates separable convolutional blocks with the U-Net architecture to effectively capture contextual features and improve pixel-level representation towards lesion extraction. It struggles with complex lesion textures and extreme interclass variations. SLT-Net [15] addresses multi-scale contextual information extraction using a multi-scale context Transformer for lesion extraction. The model faces challenges with computational efficiency and generalization to highly diverse skin lesion appearances. MHorUNet [16] utilizes recursive gated convolutions within a UNet framework for lesion extraction. The network requires high computational resources due to recursive gate convolution and multi-stage dimensional fusion. Ünver et al. [17] utilized YOLO integrated with grabcut for lesion extraction. However, the method struggles with accurately segmenting lesions with highly indistinct boundaries or severe artifacts. SapFormer is a self-adaptive position-aware skin lesion segmentation model proposed by Xin et al. [18] that can adjust to various positional parameters and capture fine-grained information, global context, and spatial connections. Its applicability is limited in contexts with limited resources and high computational complexity. The paper [19] proposes MD-UNet,

integrating tokenized MLP modules along with attention mechanism and Inception modules. However, it suffers from inaccurate edge segmentation and unsmooth edges in segmentation results. SUTrans-NET [20] uses a dual-encoder network with a transformer-based CNN for capturing the local and global contextual information. But it has limitations in segmenting skin lesions in cases where the lesion's edges are not clearly visible against the background. Liu et al. [21] proposes a multi-scale feature fusion architecture for the extraction of the skin lesions, incorporating an Efficient Channel Attention mechanism and Dense Atrous Spatial Pyramid Pooling module to enhance the model's focus on key regions and capture multi-level semantic information. However, it suffers from high computational complexity both in space and time. This work [22] proposes a method for enhancing multi-class skin lesion prediction using feature importance assessment. However, it suffers from limited generality due to the usage of limited data. The work in [23] utilized a stacked sparse autoencoder model for optimal automated skin lesion detection and classification. However, the usage of pre-processing techniques makes it susceptible to increased complexity. Zou et al. [24] proposed a DEU-Net network capturing local features and global features but suffers from accurately segmenting skin lesions with high irregularities and low contrast. Cheng et al. [25] proposed the EA-Net model, which integrates pixel-level and spatial multi-scale attention mechanisms for lesion extraction. However, it is incapable of handling class imbalance issues. MEFP-Net [26] achieves better performance than its counterparts but struggles with high computational complexity along with being incapable of extracting lesions with blurred boundaries. Zou et al. [27] introduced a GAN-based hybrid network considering both local and global features. Though the approach has better performance than its counterparts, it lacks generalizability due to the absence of diverse skin tones in the training data.

From the previous works discussed, it may be deduced that there is a need for a lightweight network that should work in a constraint scenario with reduced complexity for skin lesion extraction. Moreover, a real-time CAD system utilizing the deep learning framework is said to be effective if it is trained on fewer dermoscopic frames and can extract lesions from diverse unseen dermoscopic images, which may be used in real-time applications. We have addressed this substantially by designing a lightweight DeepSkinNet network that utilizes an encoder-decoder-based framework for the effective preservation of skin lesions from input scans.

3 DEVELOPED METHOD

3.1 DeepSkinNet

The proposed DeepSkinNet architecture comprises an encoder network, a DIL module, and a decoder network. The encoder network can preserve low, middle and high-level features accurately. The DIL module consists of a dilated convolution layer, an instance normalization layer, and a Leaky-ReLU layer, which can retain dense features accurately. The decoder framework up-samples the features efficiently with reduced noise. The detailed architecture of the proposed framework is shown in Figure 1.

Encoder architecture. The encoder architecture consists of three blocks, E1, E2, and E3. The E1 block comprises two convolutional layers (kernel = 3×3 and 64 filters), two ReLU functions, and a maxpooling layer. The convolutional layers apply filters to extract local patterns, such as textures, borders, and shapes of skin lesions, capturing important visual information while reducing computational complexity

through parameter sharing. The ReLU activation function ensures that the model can learn complex patterns and improve feature differentiation by eliminating negative values. Together, these layers enable the network to segment skin lesions, enhancing the accuracy and robustness of the segmentation process. Further, the max-pooling layer can reduce the spatial dimensions of the input features through the retention of the most important features while discarding irrelevant details. By selecting the maximum value from small regions with a size of 2×2, max-pooling helps to down-sample the image, making the model more computationally efficient. Let β_1 is the output feature map produced from the max-pooling layer after the E1 block, then it can be expressed as $\beta_1 \in \Re^{112 \times 112 \times 64}$.

Also, the E2 block comprises two convolutional layers with a kernel size of 3×3 and 128 filters, two ReLU functions, and a max-pooling layer. The E2 block can preserve the middle-level features such as contours and boundaries of the lesion, distinctive texture patterns (such as color variations or surface irregularities), and regions of interest such as darker or lighter spots that may indicate key diagnostic features. Let β_2 is the output feature map produced from the max-pooling layer after the E2 block, then it can be expressed as $\beta_2 \in \Re^{56 \times 56 \times 128}$.

Furthermore, the E3 block consists of two convolutional layers with 3×3 kernels and 256 filters, two ReLU functions, a dropout layer, and a max-pooling layer. The E3 block can preserve the middle-level features, such as more abstract and complex characteristics that describe the overall structure and nature of the lesion. By randomly deactivating a portion of neurons during training, the dropout layer avoids overfitting. This forces the model to learn more robust and generalized features, as it cannot rely on any single neuron or connection. If β_3 is the output feature map produced from the max-pooling layer after the E3 block, then it can be expressed as $\beta_3 \in \Re^{28 \times 28 \times 256}$.

Dropout enhances the model's capacity to generalize over a range of lesion sizes, forms, and textures in the setting of skin lesion extraction, improving the performance of the designed framework. In this work, the dropout rate is chosen as 0.5. For input of size $224 \times 224 \times 1$, the encoder network produces in-depth features of size $28 \times 28 \times 256$.

DIL module. The contextual relationships among pixels in skin lesion segmentation are crucial for capturing the broader structure and patterns within the image. By considering neighboring pixels, the model can better understand the spatial arrangement and texture of the lesion, improving its ability to distinguish the lesion from surrounding tissue. These relationships help in identifying the lesion's boundaries, irregularities, and more subtle characteristics, enhancing segmentation accuracy. In skin lesion segmentation, understanding context reduces errors caused by noise and allows the model to maintain consistency across regions, leading to more precise and reliable lesion delineation, even in challenging cases. Therefore, in this paper, the DIL module was developed, comprising three branches, where each branch consists of a dilated convolutional layer (sampling rate of 2, 3×3 kernels, and 32 filters) and instance normalization followed by the Leaky-ReLU activation function. The dilated convolutions expand the receptive field without increasing computational cost, allowing the model to capture more context and global information, which is essential for identifying lesions with irregular shapes or varying textures. Instance normalization helps standardize each feature map across different images, improving consistency and reducing variability caused by differences in lighting or contrast, which is common in skin lesions. Leaky ReLU, as an activation function, prevents the "dying ReLU" problem by allowing small negative values, which helps maintain gradient flow between the encoder and the decoder networks.

Together, these techniques enhance the model's ability to accurately segment skin lesions, especially in complex data. Each branch of the DIL module produces the features of size $28 \times 28 \times 32$ for input features of size $28 \times 28 \times 256$. For better feature representation, features of all three branches are concatenated in a channel that produces features of size $28 \times 28 \times 96$. If λ is the output produced from the DIL module, then it can be represented as $\lambda \in \Re^{28 \times 28 \times 96}$.

Decoder architecture. The proposed decoder network consists of three blocks as D1, D2, and D3. The D1 block comprises of two convolutional layers with 3×3 kernels and 512 filters, three ReLU functions, a transposed convolution layer with a filter size of 2×2, 256 filters, and a dropout layer. The convolutional layers extract hierarchical features while reducing noise, the ReLU activation introduces non-linearity to enhance feature learning, and the transposed convolution layers up-sample the feature maps, recovering spatial resolution lost due to down-sampling at the encoder side. This combination allows the decoder to refine boundaries preserve lesion structures that are suitable for skin lesion extraction. Further, the use of the dropout layer (dropout rate = 0.5) reduces overfitting as well as the training time. The D1 block produces the features of size $56 \times 56 \times 256$ for the input of dimension $28 \times 28 \times 96$. For better feature representation, a feature fusion mechanism is introduced after each decoder block. The output of the D1 block is combined with the output produced from the second ReLU activation layer of E3 block. The expression for the fused feature is expressed in equation (1).

$$ff_1 = (\beta_3 \oplus \gamma_1) \in \Re^{56 \times 56 \times 256} \tag{1}$$

where, γ_1 is the output produced from decoder block D1.

Furthermore, the fused feature maps are given to the D2 block of the decoder network, which consists of two convolutional layers (kernel dimension = 3×3 and 256 filters), three ReLU functions, and a transposed convolution layer with a filter size of 2×2, 128 filters. The D2 block produces the features of size $112 \times 112 \times 128$ for the input of dimension $56 \times 56 \times 256$. The output of the D2 block is combined with output produced from the second ReLU activation layer of E2 block. The expression for the fused feature after the D2 block is expressed in equation (2).

$$ff_2 = (\beta_2 \oplus \gamma_2) \in \Re^{112 \times 112 \times 128} \tag{2}$$

Again, the fused features are passed through D3 block which stacks two convolutional layers with 3×3 kernels and 128 filters, three ReLU functions, a transposed convolution layer with filter size of 2×2, 64 filters. The outcomes from D3 block with dimension $224 \times 224 \times 64$ are added with the output produced from the second ReLU activation layer of E1 block as well as the input image through a convolutional layer of 64 filters with size 3×3 and batch normalization layer followed by a LeakyReLU activation. The expression for the fused feature after the D3 block is expressed in equation (3).

$$ff_3 = (\beta_1 \oplus \gamma_3 \oplus c) \in \Re^{224 \times 224 \times 64} \tag{3}$$

where, γ_3 is the output produced from decoder block D3, and c is the output produced from the LeakyReLU activation function.

Classification block. The proposed classification block, which hybridizes three convolutional layers with the first two layers having 64 filters and the third layer having two filters (all with a 3×3 filter size), combined with two ReLU activations

and a softmax classifier. The first two convolutional layers with 64 filters effectively capture complex spatial features and hierarchical patterns, while the smaller filter size (3 × 3) ensures precise feature extraction and reduced computational overhead. The third layer with 2 filters serves as a bottleneck to reduce dimensionality and focus on two class classification: lesion and background. The ReLU activations introduce non-linearity to enhance feature learning and prevent vanishing gradients, while the softmax classifier ensures probabilistic interpretation, enabling robust and interpretable classification performance, particularly in tasks such as skin lesion segmentation where fine-grained discrimination is critical.

4 DATASETS

The work is validated through different benchmark datasets: ISIC 2016 [28], ISIC 2017 [29], ISIC 2018 [30], and PH^2 [31]. The ISIC 2016 dataset contains 900 images of varying resolutions with their respective groundtruths. The ISIC 2017 database has 2000 dermoscopic scans, which have both the melanocytic and benign types. Similarly, the ISIC 2018 dataset consists of 10,015 images of diverse clinical settings with seven different lesion types. The ISIC datasets image resolution vary from 700 × 600 to 1024 × 1024. The PH^2 dataset has 200 scans of dimension 768 × 560. However, the PH^2 emphasizes high-quality dermoscopic images for melanocytic analysis. All the images are available in .jpeg format.

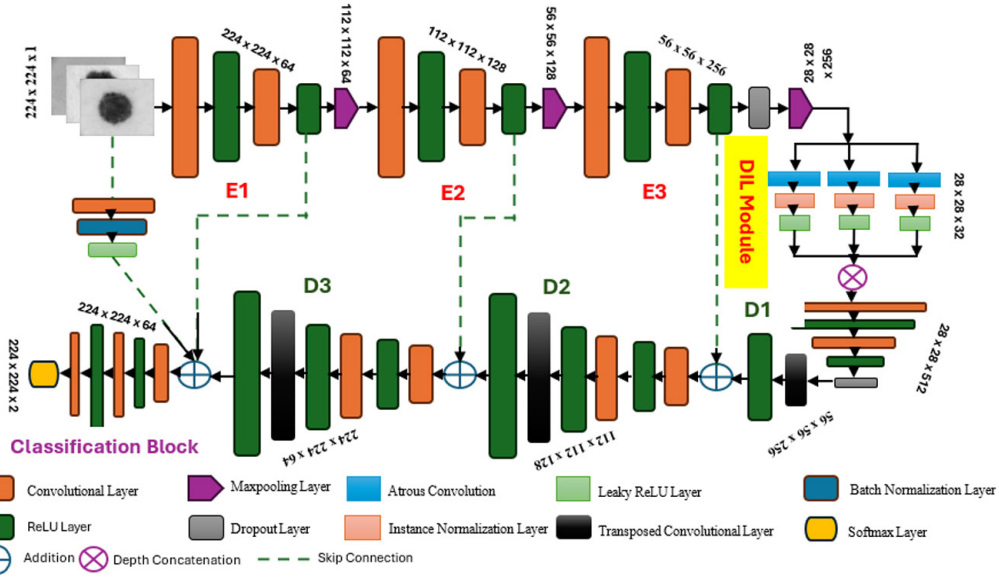

Fig. 1. Detailed framework of the designed network

5 OBJECTIVE AND SUBJECTIVE EVALUATION

The objective and subjective evaluation of the designed DILEDNet network is done utilizing benchmarked dataset including ISIC 2016 [28], ISIC 2017 [29], ISIC 2018 [30], and PH^2 [31]. In objective evaluation, the popularly used performance metrics such as accuracy (Acc) [32], specificity (Spe) [33], sensitivity (Sen) [34], and Dice Coefficient (DC) [35] were considered. The developed model is trained and tested on a Core i3 computer equipped with 16GB of RAM. MATLAB programming

using the Windows 11 operating system is used to implement the model. A batch size of 16 and 50 epochs are used to train the developed model. The developed algorithm uses the ADAM optimizer, which has a learning rate of 0.001, for adaptable and quick learning. Model convergence and effective training are balanced by using 50 epochs, a batch size of 16, and a learning rate of 0.001.

In this work, 80% of the samples are used for training, and 20% of the samples are used for validation. However, all the frames of the datasets are used for testing purposes. Here, we have adopted the Adam optimizer with a learning rate of 0.0001. Since the adaptable learning rate is one benefit in terms of adaptive learning rate and noisy gradients for which it has been efficient in the noise-laden gradients incurred with medical images, Adam is beneficial specifically for skin lesion delineation. Adam tends to combine the benefits of momentum and adaptive optimization so that convergence in the fastest possible manner is always ensured and robust performance even when different datasets are used. Sparse gradients are easy to manage with Adam so that localized lesion features are nicely elucidated. The maximum epoch considered for experimentation is 80 with a batch size of 16. In this work, the proposed DeepSkinNet is trained without any preprocessing strategies.

Table 1. Performance of the developed approach against SOTA for ISIC 2016

Approaches	Acc (%)	Spe (%)	Sen (%)	DC (%)
ARU-GD [36] 2022	94.38	94.65	89.86	90.83
BCDU-Net [37] 2019	91.78	96.20	78.11	80.95
CPFNet [38] 2020	95.09	95.91	92.11	90.23
DAGAN [39] 2020	95.82	95.68	92.28	90.85
ETV-Net [40] 2022	94.22	95.13	93.08	92.35
UNETR [41] 2022	81.45	93.27	62.91	68.83
CFPNet-M [42] 2023	95.08	95.66	92.42	94.10
Dynamic U-Net [43] 2024	95.20	96.71	**93.29**	**93.41**
FAT-Net [13] 2022	96.04	96.02	92.59	91.59
Separable-Unet [14] 2019	95.67	94.68	93.14	89.95
Swin-Unet [44] 2022	96.00	95.79	92.27	88.94
Proposed DeepSkinNet	**96.33**	**97.96**	92.84	93.33

Table 2. Performance of the developed approach against SOTA for ISIC 2017

Approaches	Acc (%)	Spe (%)	Sen (%)	DC (%)
ARU-GD [36] 2022	93.88	96.31	88.31	87.89
BCDU-Net [37] 2019	91.63	97.09	76.46	78.11
DAGAN [39] 2020	93.26	97.25	83.63	84.25
ETV-Net [40] 2022	93.38	95.58	85.78	87.66
UNETR [41] 2022	86.65	89.67	72.84	60.95
CFPNet-M [42] 2023	93.21	89.98	82.55	87.12
FAT-Net [13] 2022	93.26	97.25	83.92	85.00
SLT-Net [15] 2022	–	**97.27**	73.63	67.90
Proposed DeepSkinNet	**94.20**	92.43	**91.83**	**89.00**

Table 3. Performance of the developed approach against SOTA for ISIC 2018

Approaches	Acc (%)	Spe (%)	Sen (%)	DC (%)
ARU-GD [36] 2022	94.23	**96.81**	91.42	89.16
BCDU-Net [37] 2019	93.70	98.20	78.50	85.10
DAGAN [39] 2020	93.24	95.88	90.72	88.07
ETV-Net [40] 2022	91.02	90.60	83.79	86.37
UNETR [41] 2022	91.14	92.86	70.21	85.05
CFPNet-M [42] 2023	90.13	90.13	85.20	85.43
Dynamic U-Net [43] 2024	90.32	91.35	85.63	86.96
SLT-Net [15] 2022	–	99.35	78.85	82.54
MHorU-Net [16] 2024	91.24	92.34	86.90	87.28
Proposed DeepSkinNet	**95.53**	96.78	**92.40**	**92.05**

Table 4. Performance of the developed approach against SOTA for PH^2

Approaches	Acc (%)	Spe (%)	Sen (%)	DC (%)
YOLO2+GrabCut [17] 2019	92.99	94.02	83.63	88.13
CDNN [45] 2017	93.40	97.50	82.50	84.90
DeepCNN [46] 2015	93.20	97.80	82.00	84.70
ResNets [47] 2017	93.40	**98.50**	80.20	84.40
DeepFRCNN [48] 2018	**94.03**	96.69	85.40	87.08
Proposed DeepSkinNet	93.61	92.27	**95.18**	**91.24**

To test the performance of the designed DeepSkinNet, we have compared its results with various recently developed SOTA approaches and reported in Tables 1–4 for challenging datasets. In them Tables 1–4, the bold value indicates the best values for the said performance metric. From Table 1, it may be noticed that the developed DeepSkinNet outperforms all other methods, achieving the highest Acc (96.33%) and Spe (97.96%), along with a strong DC (93.33%) on the ISIC 2016 dataset. While models such as FAT-Net and Dynamic U-Net also show competitive performance, UNETR lags significantly with lower Acc (81.45%) and Sen (62.91%). Overall, DeepSkinNet demonstrates superior segmentation capability, which indicates that the designed framework effectively extracts the lesion from the dermoscopic images. Similarly, from Table 2, it is observed that the suggested DeepSkinNet attains the highest Sen (91.83%) and a strong DC (89.00%), showing superior segmentation performance on ISIC 2017 dataset. While SLT-Net has the highest Spe (97.27%), its Sen (73.63%) and DC (67.90%) are significantly lower, making it less balanced. As a result, the designed DeepSkinNet model provides a better trade-off between specificity and sensitivity, making it the most effective segmentation model. Further, Table 3 shows the designed DeepSkinNet acquires the highest Acc (95.53%), Sen (92.40%), and DC (92.05%), making it the most effective model overall on ISIC 2018. While SLT-Net has the highest Spe (99.35%), its lower Sen (78.85%) and DC (82.54%) suggest it struggles with balanced segmentation. Therefore, the developed DeepSkinNet provides the best trade-off across all measures, ensuring better performance. Furthermore, the efficacy of the developed algorithm is validated in the challenging PH^2 dataset. Table 4 indicates that the The suggested DeepSkinNet attains the highest Sen (95.18%) and DC (91.24%), demonstrating outstanding performance in detecting

positive cases. While models such as ResNets and DeepCNN have higher Spe (98.50% and 97.80%, respectively), their lower Sen (80.20% and 82.00%) indicates a weaker ability to detect true positives. From, this analysis, it is concluded that the designed algorithm can extract the lesion from the challenging datasets accurately with reduced misclassification of healthy skin regions.

The visual demonstration of the developed DeepSkinNet model is shown in Figure 2. Figures 2a and 2b represent the original dermoscopic image and its green channel respectively. Figure 2c indicates the ground truth image in their respective dataset. Similarly, Figures 2d and 2e represent the semantic segmentation outcomes of the designed DeepSkinNet and the corresponding segmentation mask respectively. From Figure 2e it is found that the proposed technique is capable of extracting fine as well as coarse-scale lesion features efficiently. However, by comparing the segmentation masks provided in Figure 2e with the ground truth masks shown in Figure 2c, it may be noticed that the designed DeepSkinNet effectively delineates the lesion boundary but it incorrectly classifies some fine foreground regions (in this case it is the lesion details) as healthy skin (in this case it is background). Further, to validate the efficacy of the proposed DeepSkinNet, the computational complexity of the model is explained in Table 5 in terms of space complexity and computational time. From Table 5 it can be seen that the computational time of the proposed model is better for the ISIC 2016 dataset as compared to other benchmark datasets.

Table 5. Computational complexity of the DeepSkinNet for different datasets

Datasets	Space Requirement	Computational Time
ISIC 2016	2.7 M	0.025 sec
ISIC 2016	2.7 M	0.031 sec
ISIC 2016	2.7 M	0.041 sec
PH^2	2.7 M	0.027 sec

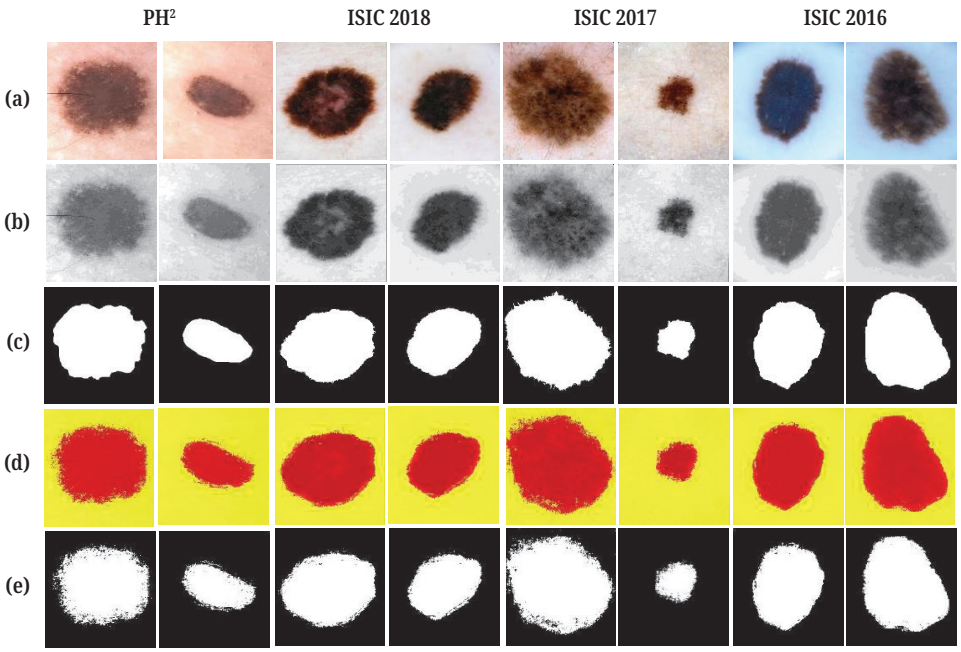

Fig. 2. Visual demonstration of the proposed DeepSkinNet model on skin lesion datasets: (a) input scan, (b) green channel scan, (c) ground truth provided from the dataset, (d) semantic segmentation output, and (e) extracted mask

6 CONCLUSION

In this study, a unique and lightweight DeepSkinNet is developed for skin lesion extraction from dermoscopic images. The proposed DeepSkinNet model comprises an encoder framework, a DIL module, and a decoder network. The encoder network is capable of handling various artifacts and providing the in-depth features of the lesion regions effectively. The designed DIL module can preserve the subtle details of the lesion region of varying sizes and structures. The proposed decoder framework learns a mapping from in-depth feature level to pixel level. The effectiveness of the developed DeepSkinNet model is validated through various benchmarked datasets such as PH^2, ISIC 2017, ISIC 2018, and ISIC 2016. Also, the efficiency of the developed DeepSkinNet model is tested on the unseen setup and found to provide better accuracy, which may be suitable for real-time applications. Further, the designed model shows its superiority in terms of various quantitative measures and surpasses 18 recently developed SOTA approaches.

7 REFERENCES

[1] M. Adamu Mohammed, O. Bismark, S. Alornyo, M. Asante, and B. Obo Essah, "ResFCNET: A skin lesion segmentation method based on a deep residual fully convolutional neural network," *IETI Trans. Data Anal. Forecast.*, vol. 1, no. 1, pp. 4–19, 2023. https://doi.org/10.3991/itdaf.v1i1.35723

[2] S. Pathan, K. G. Prabhu, and 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. https://doi.org/10.1016/j.bspc.2017.07.010

[3] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020. https://doi.org/10.1109/ACCESS.2019.2960504

[4] R. Ray, S. Jena, P. Parida, L. Dash, and S. K. Biswal, "Empowering diabetic eye disease detection: Leveraging differential evolution for optimized convolution neural networks," *Int. J. Online Biomed. Eng.*, vol. 20, no. 10, pp. 86–100, 2024. https://doi.org/10.3991/ijoe.v20i10.49187

[5] W. Tang and J. Wang, "A review of joint applications of IoT and deep learning," *IETI Trans. Data Anal. Forecast.*, vol. 1, no. 3, pp. 4–17, 2023. https://doi.org/10.3991/itdaf.v1i3.44517

[6] I. D. V. Garcia Carreño, "e-Leadership: A bibliometric analysis," *Int. J. Adv. Corp. Learn.*, vol. 13, no. 1, pp. 19–34, 2020. https://doi.org/10.3991/ijac.v13i1.12341

[7] J. Kawahara and G. Hamarneh, "Multi-resolution-Tract CNN with hybrid pretrained and skin-lesion trained layers," in *Machine Learning in Medical Imaging. MLMI 2016*, in Lecture Notes in Computer Science (LNIP), L. Wang, E. Adeli, Q. Wang, Y. Shi, and H. I. Suk, Eds., vol. 10019, 2016, pp. 164–171. https://doi.org/10.1007/978-3-319-47157-0_20

[8] K. M. Sudar, P. Nagaraj, V. Muneeswaran, B. Panda, and A. K. Bhoi, "Dermo classify: A dermatologist skin disease detection and classification using DCNN," *Res. Biomed. Eng.*, vol. 41, 2025. https://doi.org/10.1007/s42600-024-00392-1

[9] J. D. Dorathi Jayaseeli *et al.*, "An intelligent framework for skin cancer detection and classification using fusion of Squeeze-Excitation-DenseNet with Metaheuristic-driven ensemble deep learning models," *Sci. Rep.*, vol. 15, 2025. https://doi.org/10.1038/s41598-025-92293-1

[10] A. M. Vidhyalakshmi and M. Kanchana, "Optimizing skin cancer diagnosis: A modified ensemble convolutional neural network for classification," *Microsc. Res. Tech.*, vol. 88, no. 6, pp. 1681–1700, 2025. https://doi.org/10.1002/jemt.24792

[11] P. Radhakrishnan and P. K. Sukumar, "Optimized attention transfer for robust and scalable dermatology image classification," *Microsc. Res. Tech.*, 2025. https://doi.org/10.1002/jemt.24823

[12] E. H. I. Eliwa, "Enhancing skin cancer diagnosis through fine-tuning of pretrained models: A two-phase transfer learning approach," *Int. J. Breast Cancer*, vol. 2025, no. 1, p. 4362941, 2025. https://doi.org/10.1155/ijbc/4362941

[13] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, "FAT-Net: Feature adaptive transformers for automated skin lesion segmentation," *Med. Image Anal.*, vol. 76, p. 102327, 2022. https://doi.org/10.1016/j.media.2021.102327

[14] P. Tang *et al.*, "Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging," *Comput. Methods Programs Biomed.*, vol. 178, pp. 289–301, 2019. https://doi.org/10.1016/j.cmpb.2019.07.005

[15] K. Feng, L. Ren, G. Wang, H. Wang, and Y. Li, "SLT-Net: A codec network for skin lesion segmentation," *Comput. Biol. Med.*, vol. 148, p. 105942, 2022. https://doi.org/10.1016/j.compbiomed.2022.105942

[16] R. Wu *et al.*, "MHorUNet: High-order spatial interaction UNet for skin lesion segmentation," *Biomed. Signal Process Control*, vol. 88, p. 105517, 2024. https://doi.org/10.1016/j.bspc.2023.105517

[17] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, p. 72, 2019. https://doi.org/10.3390/diagnostics9030072

[18] C. Xin *et al.*, "Transformer guided self-adaptive network for multi-scale skin lesion image segmentation," *Comput. Biol. Med.*, vol. 169, p. 107846, 2024. https://doi.org/10.1016/j.compbiomed.2023.107846

[19] Y. Chen, X. Sun, Y. Duan, Y. Wang, J. Zhang, and Y. Zhu, "Lightweight semantic segmentation network for tumor cell nuclei and skin lesion," *Front. Oncol.*, vol. 14, 2024. https://doi.org/10.3389/fonc.2024.1254705

[20] Y. Li, T. Tian, J. Hu, and C. Yuan, "SUTrans-NET: A hybrid transformer approach to skin lesion segmentation," *PeerJ Comput. Sci.*, vol. 10, p. e1935, 2024. https://doi.org/10.7717/peerj-cs.1935

[21] L. Liu, X. Zhang, Y. Li, and Z. Xu, "An improved multi-scale feature fusion for skin lesion segmentation," *Appl. Sci.*, vol. 13, no. 14, p. 8512, 2023. https://doi.org/10.3390/app13148512

[22] A. Paulauskaite-Taraseviciene, K. Sutiene, N. Dimsa, and S. Valiukeviciene, "Enhancing multi-class prediction of skin lesions with feature importance assessment," *Int. J. Appl. Math. Comput. Sci.*, vol. 34, no. 4, 2024. https://doi.org/10.61822/amcs-2024-0041

[23] K. A. Ogudo, R. Surendran, and O. Ibrahim Khalaf, "Optimal artificial intelligence based automated skin lesion detection and classification model," *Comput. Syst. Sci. Eng.*, vol. 44, no. 1, pp. 693–707, 2023. https://doi.org/10.32604/csse.2023.024154

[24] A. Karimi, K. Faez, and S. Nazari, "DEU-Net: Dual-Encoder U-Net for automated skin lesion segmentation," *IEEE Access*, vol. 11, pp. 134804–134821, 2023. https://doi.org/10.1109/ACCESS.2023.3337528

[25] D. Cheng *et al.*, "EA-Net: Research on skin lesion segmentation method based on U-Net," *Heliyon*, vol. 9, no. 12, p. e22663, 2023. https://doi.org/10.1016/j.heliyon.2023.e22663

[26] S. Hao *et al.*, "MEFP-Net: A dual-encoding multi-scale edge feature perception network for skin lesion segmentation," *IEEE Access*, vol. 12, pp. 140039–140052, 2024. https://doi.org/10.1109/ACCESS.2024.3467678

[27] R. Zou, J. Zhang, and Y. Wu, "Skin lesion segmentation through generative adversarial networks with global and local semantic feature awareness," *Electronics*, vol. 13, no. 19, p. 3853, 2024. https://doi.org/10.3390/electronics13193853

[28] D. Gutman *et al.*, "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*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1605.01397

[29] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168–172. https://doi.org/10.1109/ISBI.2018.8363547

[30] N. Codella *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," *arXiv preprint arXiv.1902.03368*, 2019. https://doi.org/10.48550/arXiv.1902.03368

[31] C. Barata, M. Ruela, M. Francisco, T. Mendonca, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Syst. J.*, vol. 8, no. 3, pp. 965–979, 2014. https://doi.org/10.1109/JSYST.2013.2271540

[32] R. Rout, P. Parida, M. K. Panda, and S. S. Rout, "Saliency induced fusion for skin lesion detection," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 16, no. 3, pp. 696–710, 2024.

[33] R. Rout, P. Parida, and S. Dash, "Automatic skin lesion segmentation using a hybrid deep learning network," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 15, pp. 238–249, 2023.

[34] R. Rout, P. Parida, and S. Dash, "A hybrid deep learning network for skin lesion extraction," in *Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022). SoCPaR 2022*, in Lecture Notes in Networks and Systems, A. Abraham, T. Hanne, N. Gandhi, P. Manghirmalani Mishra, A. Bajaj, and P. Siarry, Eds., Springer, Charm., vol. 648, pp. 682–689, 2023. https://doi.org/10.1007/978-3-031-27524-1_66

[35] R. Rout, P. Parida, and S. Patnaik, "Melanocytic skin lesion extraction using mean shift clustering," in *2021 International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA)*, 2021, pp. 565–574. https://doi.org/10.1109/ICEITSA54226.2021.00112

[36] D. Maji, P. Sigedar, and M. Singh, "Attention Res-UNet with guided decoder for semantic segmentation of brain tumors," *Biomed. Signal Process. Control*, vol. 71, p. 103077, 2022. https://doi.org/10.1016/j.bspc.2021.103077

[37] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-Directional ConvLSTM U-Net with densley connected convolutions," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 406–415. https://doi.org/10.1109/ICCVW.2019.00052

[38] S. Feng *et al.*, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 10, pp. 3008–3018, 2020. https://doi.org/10.1109/TMI.2020.2983721

[39] B. Lei *et al.*, "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Med. Image Anal.*, vol. 64, p. 101716, 2020. https://doi.org/10.1016/j.media.2020.101716

[40] M. Nour, H. Öcal, A. Alhudhaif, and K. Polat, "Skin lesion segmentation based on edge attention Vnet with balanced focal Tversky loss," *Math. Probl. Eng.*, vol. 2022, no. 1, p. 4677044, 2022. https://doi.org/10.1155/2022/4677044

[41] A. Hatamizadeh *et al.*, "UNETR: Transformers for 3D medical image segmentation," arXiv preprint arXiv.2103.10504, 2021. [Online]. Available: http://arxiv.org/abs/2103.10504

[42] A. Lou, S. Guan, and M. Loew, "CFPNet-M: A light-weight encoder-decoder based network for multimodal biomedical image real-time segmentation," *Comput. Biol. Med.*, vol. 154, p. 106579, 2023. https://doi.org/10.1016/j.compbiomed.2023.106579

[43] J. Yang, D. S. Marcus, and A. Sotiras, "Dynamic U-Net: Adaptively calibrate features for abdominal multi-organ segmentation," arXiv preprint arXiv.2403.07303, 2024. [Online]. Available: http://arxiv.org/abs/2403.07303

[44] S. Atek, I. Mehidi, D. Jabri, and D. E. C. Belkhiat, "SwinT-Unet: Hybrid architecture for medical image segmentation based on Swin transformer block and Dual-Scale Information," in *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, 2022, pp. 1–6. https://doi.org/10.1109/ISPA54004.2022.9786367

[45] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," *arXiv preprint arXiv:1703.05165*, 2017. https://doi.org/10.48550/arXiv.1703.05165

[46] C. Szegedy *et al.*, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[47] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale Dermoscopy images and deep residual networks," arXiv preprint arXiv:1703.04197, 2017. https://doi.org/10.48550/arXiv.1703.04197

[48] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, 2018. https://doi.org/10.1016/j.cmpb.2018.05.027

8 AUTHORS

Abhipsa Pattanaik is pursuing her PhD in the School of Computer Science Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India. Her study interests include Biomedical Data Analysis, Computer Vision, Biomedical Image Processing, and Artificial Intelligence (E-mail: 2181057@kiit.ac.in).

Dr. Leena Das is an Associate Professor at the School of Computer Science Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India. Her areas of interest include Parallel and Distributed Computing, Performance and Reliability Analysis of Interconnection Networks, High-Dimensional Data, Machine learning, Soft Computing, and Big Data Analytics.