RESEARCH ARTICLE

WILEY

A novel two-staged network for skin disease detection using atrous residual convolutional networks

Karthik Ramamurthy[1] | **Anirudh Muthuswamy**[2] | **Nagharjun Mathimariappan**[2] | **Gugan S. Kathiresan**[2]

[1]Centre for Cyber Physical Systems, School of Electronics Engineering, Vellore Institute of Technology, Chennai, India

[2]School of Electronics Engineering, Vellore Institute of Technology, Chennai, India

Correspondence
Karthik Ramamurthy, Centre for Cyber Physical Systems, School of Electronics Engineering, Vellore Institute of Technology, Chennai, India.
Email: r.karthik@vit.ac.in

Summary

The preferred modality for the identification of skin disease has predominantly been dermoscopy. Skin lesion identification is the primary diagnostic basis for the dermatologist to evaluate the severity and impact of the disease. Current diagnostic procedures for the identification of skin lesions may be subjected to misdiagnosis among observers. Also, these procedures are conventionally labour-intensive and prone to delays in treatment. Deep learning has proven to be beneficial for automated computer-aided diagnosis in the medical field. This research presents a two-stage approach involving segmentation and classification architectures for the effective detection of skin disease. The proposed approach aims to combine the architectural benefits of residual learning with the contextual retention of atrous convolutions. The classification is performed on seven different classes of skin lesions namely, melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis/Bowen's disease (AKIEC), benign keratosis (BKL), dermatofibroma (DF), and vascular lesion (VASC). To the best of our knowledge, this is the first research that reports the impact of a two-staged detection network using atrous residual convolutions for these seven classes. Experimental observations indicate that the proposed model is characteristically balanced in terms of inter-class performance, with precise segmentation. The proposed network was trained with the HAM10000 dataset and reported a classification accuracy and precision of 89.27% and 89.06%, respectively.

KEYWORDS

atrous residual networks, convolutional neural networks, dermoscopy, skin lesion

1 | INTRODUCTION

A study conducted by the Indian Council of Medical Research (ICMR) in the year 2021, reported that over recent years, vulnerability to skin cancer has increased in people belonging to specific regions of South Asian developing countries.[1] One of the causes of skin cancer is attributed to occupations that spend more time in the outdoor environment.[2] Skin cancer was considered the most common form of cancer in the United States of America.[3] It has also been estimated that nonmelanoma skin cancer is the leading cancer in terms of cases among men in Oceania, Australia, and North America and among women in Oceania and Australia.[4] While skin cancer is characteristically harmful to European skin profiles, recent studies report increasing cases and severity of skin cancer in other parts of South Asian developing countries as well.

Skin cancer can be broadly classified into two types, nonmelanoma skin cancer (NMSC), and melanoma skin cancer (MSC). MSC is comparatively less prevalent in developing countries than the common NMSC. However, the health risk and mortality rate in MSC is much higher.[3] It is expected

that, the deaths caused by melanoma is expected to increase by 6.5% from 2021 to 2022.[5] The causes behind skin cancer vary in source, but are common in terms of its composition of radioactive and radiation compounds. Ultraviolet radiation, plutonium, arsenic, and cigarettes are considered as group 1 skin cancer-causing agents. Agricultural workers are susceptible to compounds like arsenic due to their presence in pesticides used regularly on farms.[2] This leaves farmers who work for long hours in the sun and are exposed to pesticides at a higher risk.

Detection of skin cancer in its preliminary stages is of vital importance. Early detection of melanoma indicates a positive 5-year survival rate is 99%.[4] Classical techniques for detecting skin cancer involve visual and physical examination of the skin by a dermatologist or other healthcare provider. During the exam, the provider looks for any suspicious growths, moles, or lesions on the skin, noting their characteristics such as size, shape, color, and texture. Dermatoscopy, a handheld device with a magnifying lens and light source, may be used to examine the skin in more detail. If a suspicious area is identified, a biopsy may be performed to remove a sample of skin tissue for microscopic examination. These techniques heavily rely on the expertise and experience of the healthcare provider, and while they can be effective, they may not always detect skin cancer at an early stage.

Of the existing methods, the skin biopsy is the most recommended method to detect the presence of skin cancer cells.[6] It involves a procedure to remove skin cells and samples, and testing them under laboratory conditions for detecting the presence of skin cancer. Skin biopsies are of three types: shave biopsy, punch biopsy, and excisional biopsy. These biopsies vary in terms of the instrument used and the requirement of the epidermis, dermis, and superficial fat. If diagnosis and treatment are not started within the first 30 days the risk of death increases by 5%, and further increases by 41% after the first 119 days.[7] As a result, without an initial diagnosis of the disease, patients are more likely to experience problems such as bleeding, bruising, scarring, infection, and allergic responses.

In addition, skin biopsies are prone to human error as well. Skin biopsy pathways and wrong-site surgeries are such common errors. In the whole process of skin biopsy, it has been estimated that there are 20 steps approximately. Each of the steps has an error rate of >0. Even if each step is 99% accurate in its tasks, the overall accuracy comes down to 82%.[8] With these risks and delays in mind, automated segmentation detection and classification of skin lesions would be pivotal to improving diagnostic performance. Also, the segmentation of skin lesions is a subjective process, as the morphological pattern of different types of lesion is similar to one another. Hence, it requires domain expertise to accurately recognize the correct class of lesion. This places a vital need to develop an automatic method for the precise segmentation of skin lesions.

In this case, leveraging the power of deep learning can potentially act as a viable diagnostic assistant, eliminating the need for the risk involved in skin biopsy. Image-based detection is a solution to both the risks involved, and can save valuable time. Thus, a computerized model can provide a much faster and more accurate prediction and detection of skin cancer. It can relieve the pressure mounted on medical officials and focus on only the essential steps in diagnosis. Accurate diagnosis of skin diseases can be performed if the skin lesion is segmented initially and then the segmented region of interest is further subjected to final classification. This will eliminate false positives as the region of interest is focused on the target region appropriately. Hence a two-staged network will serve as a proper approach for the effective diagnosis of skin diseases, where the first stage involves segmentation of skin lesions and the next stage applies a classification algorithm for finding the class label. In this research, we propose an automated two-staged deep learning framework to segment and classify skin lesions into seven different categories. It could act as a vital assistive aid for the medical professionals to precisely classify different classes of skin lesions with high precision.

2 | RELATED WORKS

The task of classifying skin lesions from dermoscopy images has been attempted in various studies. These studies have been categorized based on the employed approaches that are similar to our proposal. The majority of the works reviewed utilized real-world skin dermoscopy datasets that are concurrent with the data used in the proposed work. Section 2.1 details works that have employed classification[9–26] techniques and Section 2.2 for segmentation techniques.[27–30]

2.1 | Classification techniques

The classification approaches can be further categorized into transfer learning and pretrained approaches,[9–15] custom architectures and federated learning,[16–18] hybrid and multistage models,[19–24] and RCNN architectures.[25,26]

Transfer learning is a popular approach employed in deep learning research for its benefits in computational efficiency and performance. Modified transfer learning models have been commonly used to leverage these benefits. The MobileNet architecture is a popular implementation, as seen in studies by Anand et al.,[9] Agrahari et al.,[10] and Sae-Lim et al.[11] Calderón et al. proposed a novel bilinear CNN architecture in an attempt to improve the classification performance by two-fold.[12] The two models combine ResNet-50 and VGG-16 networks.

Residual networks are generally employed for their robustness against the vanishing gradient problem. The ResNet architecture and its variants are popular in this regard. Khan et al. employed pretrained ResNet-50 and ResNet-101 models in collaboration with support vector machine classifiers and achieved an overall classification accuracy of 89.8%.[13] Anand et al. presented an approach for the classification of skin diseases by

analyzing ResNet18, ResNet50, and ResNet101 models.[31] It was inferred that the ResNet101 model yielded optimal performance when compared to the other two versions of ResNet.

In an attempt to enhance diagnostic accuracy, Thomas et al. combined patient metadata along with the classification performance of pretrained models.[14] The GoogleNet, DenseNet-201, and Inception-Resnet-v2 architectures reported a 10% increase in accuracy when patient metadata was included in image features. In their study, Karthik et al. proposed a novel implementation of the EfficientNetV2 architecture with an attention block. This proposed Eff2Net allowed for a model with lower parameters, and improved overall performance.[15]

Deployable and explainable solutions are part of a recent trend in deep learning research. Waweru et al. employed the pretrained DenseNet-201 architecture[16] where their solutions are customized and crafted for deployment as web services or mobile applications. However, custom architectures are often employed as novel approaches to solve specific facets of a larger problem statement. Huo et al. introduced a fully deployed solution that employs a customized CNN architecture, classifying skin lesions at an accuracy of 75%.[17] Połap et al. proposed a federated learning approach that combines a range of deep learning-based classifiers.[18] These classifiers include, custom CNN architectures, and state-of-the-art models like MobileNet.

Hybrid and multistage models are often employed to harness the power of combining multiple architectures and their benefits. These combinations can span across other domains as well. For example, the study by Khan et al. introduced an IoT-based framework with multimodal skin lesion classification capabilities.[19] The model included dedicated custom segmentation and classification stages. Similarly, Mahbod reported a maximum classification accuracy of 89.8% by combining images that were preprocessed by subtracting manually created masks.[20] Khan et al. took a novel high dimension contrast transform (HDCT) based saliency segmentation stage to segment skin lesion images.[21] These segmented images were further classified by a pretrained DenseNet model stage. Cengil et al. combined paradigm machine learning classifiers with pretrained deep learning architectures.[22] An accuracy of 77.80% was reported by an AlexNet-SVM hybrid.

Fusion-based and extreme learning machine approaches make up a small niche in skin lesion classification techniques. Employing a fusion-based approach, Khan et al. proposed two stages with pretrained DenseNet and MobileNet models.[23] The extracted features were combined using a multimax coefficient correlation method and further classified using a multiclass extreme learning machine classifier. Similarly, the correlation and extreme learning techniques were also employed in the work by Khan et al., combined with an improved moth flame optimization (IMFO) algorithm.[24] The IMFO algorithm extracted the dominant features from images segmented by the deep saliency segmentation method.

RCNN architectures are effective in the bounding box and pixel-wise localization tasks with classification as well. Khan et al. utilized a Mask-RCNN architecture in their study, along with a ResNet-50 backbone and softmax classifiers.[25] The model yielded an accuracy of 86.5%. Goyal et al. proposed an ensemble approach that adds the features extracted by Mask-RCNN and DeeplabV3+ to generate a combined segmentation mask.[26] The employed ensemble reported an accuracy of 94.1%. A comparative analysis of the existing works applied for classification of skin diseases is presented in Table 1.

2.2 | Segmentation techniques

There is a wide range of segmentation techniques and approaches employed in the collected works. While these works vary in datasets used, they are common in their attempts to accurately segment skin lesions from dermoscopy images. A majority of the approaches are proposed as assistive mechanisms to doctors in real-world medical scenarios. These works can be further categorized based on the methods employed. These categories include attention-based techniques,[27,28] and pretrained or paradigm architectures.[29,30]

Sarker et al. proposed the SLSNet model aimed to be a lightweight solution designed for emergency medical applications. The model applied channel attention mechanisms to improve localization, similar to previous works. By employing a generative adversarial network (GAN) model and a multiscale aggregation mechanism, the overall effects of noise were reduced.[27] In a very similar approach with a GAN-based model that combines multiscale aggregation and 1-D channel filtering. Singh et al. proposed the FCA-Net.[28] The FCA-Net reported a better IoU score than the SLSNet,[27] however, trained with a larger dataset.

Employing pretrained and paradigm architectures allow for enhanced performance and quicker computation times. There are a range of works that customize popular architectures like UNet and UNet++. For example, Yang et al. proposed the EfficientUNet++ that combines the UNet++ architecture with an EfficientNet backbone.[29] By introducing more skip connections, the retention of information is achieved across all layers, improving overall segmentation. Another such architecture is the YOLO architecture, popularized by its single- shot learning technique. This attributes to its fast computation, suitable for real-world medical emergencies. In the model proposed by Unver et al., the YOLOv3 model was combined with the GrabCut algorithm.[30] In an overall four-stage implementation, the GrabCut algorithm improved the localization in the post-processing phase along with morphological operations as well. Anand et al. presented an automated approach for the segmentation of skin lesions using a modified U-Net architecture.[32] More kernels were added to the network for precise feature extraction. A comparative analysis of the existing works applied for segmentation of skin diseases is presented in Table 2.

TABLE 1 Comparative analysis of the existing works applied for classification of skin diseases.

S. no	Source	Method	Dataset considered	Observations
1	Anand et al.[9]	Modified MobileNet	HAM10000	Accuracy: 90% Precision: 86%
2	Agrahari et al.[10]	Pretrained MobileNet	HAM10000 ISIC	Accuracy: 80.81%
3	Sae-Lim et al.[11]	Modified MobileNet	HAM10000	Accuracy: 83.93%
4	Khan et al.[13]	ResNet 50, ResNet101 with SVM	HAM10000, ISBI 2017, and ISBI 2016	Accuracy: 89.8%
5	Karthik et al.[15]	Eff2Net	DermNet NZ, Derm7pt, DermatoWeb, DermAtlas, Hellenic Derm Atlas, Meddean	Accuracy: 84.7%
6	Połap et al.[18]	Federating learning	HAM10000	Accuracy: 89.12%
7	Khan et al.[19]	30 layered CNN	ISBI2016, ISIC2017, ISBI2018, ISIC2019, and HAM10000	Accuracy: 87.02%
8	Cengil et al.[22]	ResNet18 with SVM	HAM10000	Accuracy: 76.23%
9	Attique et al.[23]	DenseNet201	HAM10000, ISBI2018, and ISIC2019	Accuracy: 96.5%
10	Khan et al.[24]	Custom CNN	HAM10000, ISBI 2016, ISBI 2017, ISIC 2018, and PH2	Accuracy: 90.67%
11	Khan et al.[25]	24 layered CNN	HAM10000	Accuracy: 86.5%

TABLE 2 Comparative analysis of the existing works applied for segmentation of skin diseases.

S. no	Source	Method	Dataset considered	Observations
1	Sarkar et al.[27]	SLSNet	ISBI 2017 ISIC 2018	Dice: 90.63% Jaccard: 81.98%
2	Singh et al.[28]	FCA-Net	ISBI2016 ISBI2017 ISIC2018	Dice: 92.80% IoU: 86.41%
3	Yang et al.[29]	EfficientUNet++	PH2 ISIC	Dice: 93%
4	Unver et al.[30]	Yolov3 with Grabcut	PH2 ISIC 2017	Dice: 84.26% Jaccard: 74.81%
5	Anand et al.[32]	Modified U-Net	PH2	Dice: 89.01%

3 | RESEARCH GAPS AND CONTRIBUTIONS

This research aims to address the following research gaps in skin lesion localization and classification.

- The skin lesion can be categorized into multiple types. The discriminative features of each type present themselves in varying shapes and sizes. However, their structural features are quite similar. Such structural characteristics of the skin lesions increase the probability of false positives. While transfer learning methods report appreciable results in terms of the overall classification, it may lack balanced inter-class performance.

- A majority of the works employed the dataset without performing initial augmentation. However, pixel-wise augmentation has been used scarcely. This augmentation could be beneficial to improve the inter-class accuracy with a wide range of classes.

- Most of the works surveyed propose single-stage networks, that primarily address classification. These networks do not grant enough focus to the subtle morphological variations among different classes of lesions. This overlooks the potential benefits of localized focus as an input.

- Most of the existing works employ binary classification or tasks with a lesser number of classes. This is also done to ensure low-complexity models for easy deployment. More focus must be granted to detecting a wide range of classes.

The research gaps that have been identified from the literature survey serve as the basis for this work. The proposed approach looks to address these challenges, and provide solutions to it. The overall workflow of the presented approach is illustrated in Figure 1. The following are the key contributions of this research:
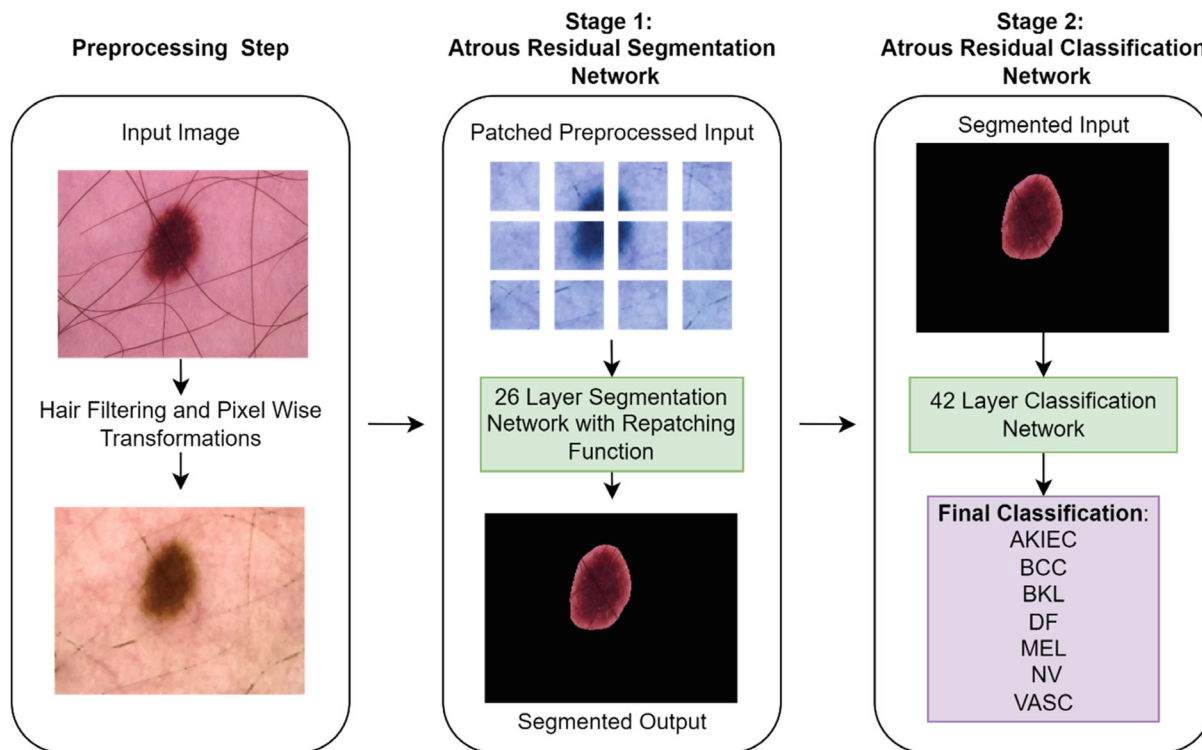
FIGURE 1 The overall workflow of the presented approach.

- This research presents a two-stage approach for the detection of skin lesions. These stages are designed to include the benefits of an atrous residual segmentation network, and a deeper modified architecture for the classification network.

- The input images are subjected to patching as the first step. Patching improves the model's efficiency in delineating the subtle variations of the lesion morphology. This was applied in the interest of boosting inter-class detection efficiency.

- To improve the overall detection performance, preprocessing steps have been employed. The initial preprocessing includes a custom combination of pixel-wise transformations and hair filtering. This allows the removal of occluding hairs and increasing the discrimination of lesion features.

- The segmentation stage employs a novel atrous residual architecture, constructed with customized layers for precise spatial feature extraction. The segmented output acts as a focused basis for further classification

- The proposed classification stage is designed with 42 layers of atrous residual convolutions. This was applied to include the benefits of a deeper network, with benchmark inter-class performance on seven types of skin disease.

4 | PROPOSED WORK

The following section details the features of the proposed approach and the motivation behind it. The overall workflow is elucidated in separate subsections.

4.1 | Dataset details

The dataset used in this research is the HAM10000 dataset from the Harvard Dataverse. The dataset consists of multisource dermoscopic images of common pigmented skin lesions. The dataset contains a total of 10,015 dermoscopic lesion images that are in JPEG format. Additionally, binary segmentation masks are provided along with the images. The dataset consists of seven classes of skin lesions, which indicate the diagnosis of each of the input lesion images.

The classes are melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis/Bowen's disease (AKIEC), benign keratosis (BKL), dermatofibroma (DF), vascular lesion (VASC).

The dataset distribution per class is presented in Figure 2. It contains details about the balanced dataset employed. While the original dataset consisted of 10,015 images, the balanced dataset was of 46,489 images. Dataset balancing was performed to improve the overall classification performance. This ensured that there is equal representation of each label in the model. Balancing was performed through data augmentation by creating multiple variations of the images using geometrical image transformations. The transformations employed were rotation, flipping, Gaussian blurring and so on. In addition, we have also applied hue saturation value modification and contrast limited adaptive histogram equalization (CLAHE) for generating more samples.

4.2 | Preprocessing

Skin lesions manifest on the visible layer of the skin. They vary in terms of their shape, color, and texture. However, the distinction between these lesions on dermoscopy images is low, due to the minimal differences in color, and similar shape. Also, dermoscopy images often have lesions occluded by the hair on the patient's skin, impacting the precision of feature extraction. This can be seen in an example in Figure 3.
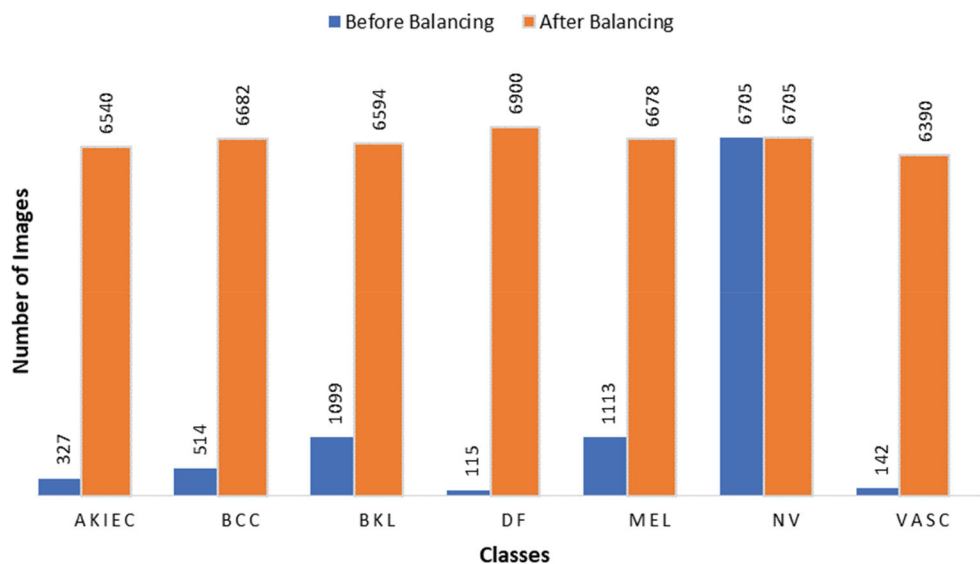
FIGURE 2 Native versus balanced dataset distribution across all classes.
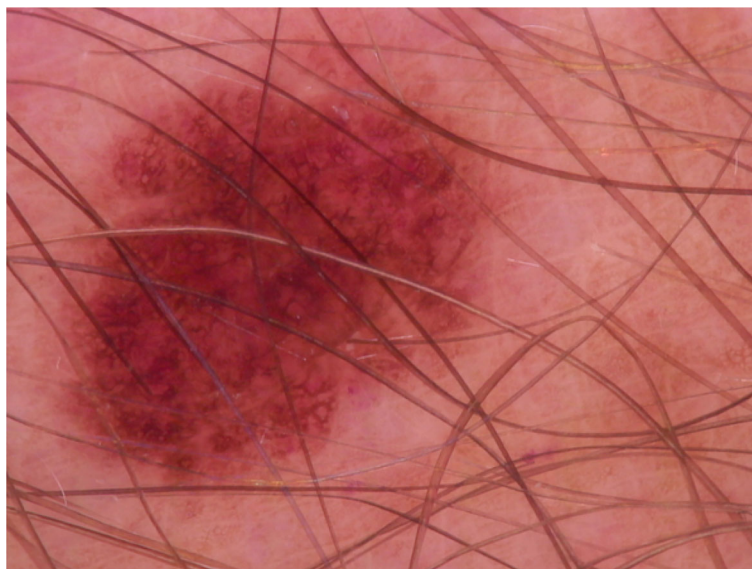
FIGURE 3 Melanocytic nevus sample with occlusion due to surface hair.

Hence, there is scope for improvement in terms of the distinctiveness of features, and the occluding hair. By applying specific preprocessing techniques to improve the distinctness of the features and to remove the occluding hair, better performance is expected.

4.2.1 | Patching of images

The native dermoscopy images exhibit varied morphological characteristics. To facilitate region-specific focus, the images are subjected to patching. Patching splits a 2-D input image into 12 patches of equal size before feeding it into the segmentation model. After the patched images undergo segmentation separately, it is repatched to obtain the final segmentation mask. Figure 4A,B represents the results of the input image before and after patching. Finally, Figure 4C is the output of the segmentation model which will be repatched to get the whole segmented image.

4.2.2 | Hair-filtering

A combination of morphological operations was applied to the native dataset. The combination included "closing" and "opening" operations with a custom filter, in the order that yielded images without occluding hairs. The relations for "closing" and "opening" operations can be found in Equations (1) and (2), respectively. An example of the proposed combination can be found in Figure 5 for the "Melanocytic Nevus" class.

$$A \cdot B = (A \bigoplus B) \ominus B, \tag{1}$$

$$A \circ B = (A \ominus B) \bigoplus B, \tag{2}$$

where $\cdot$ = closing operation, $\circ$ = opening operation, $\ominus$ = erosion, $\bigoplus$ = dilation, and A, B are the image matrices, respectively.

(A)   (B)   (C)
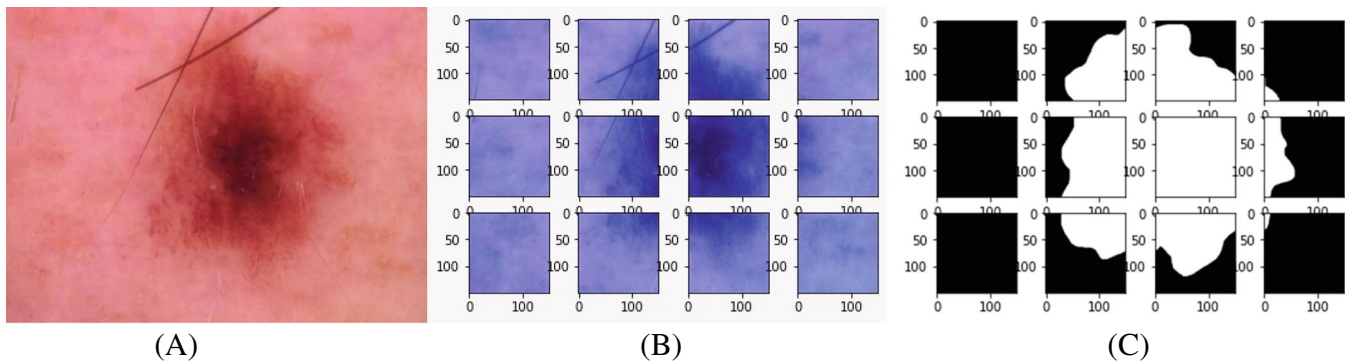
FIGURE 4 (A) Native image before patching, (B) 12-patched image with normalization, and (C) 12-patched mask after segmentation

(A)   (B)
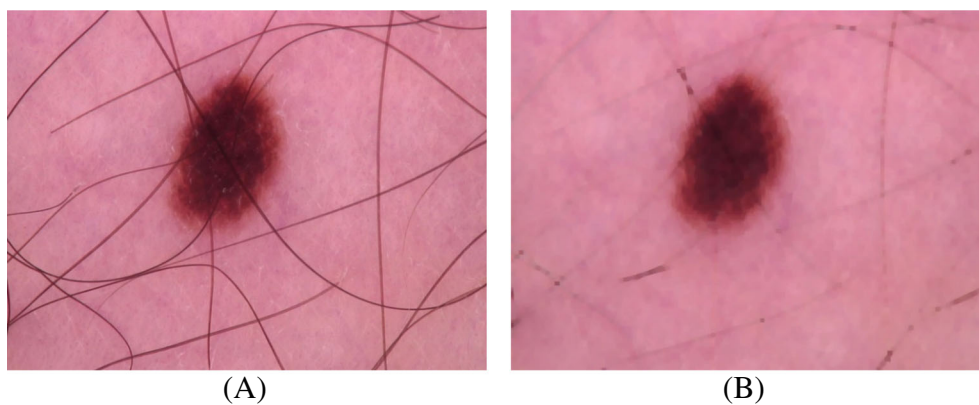
FIGURE 5 (A) Sample before hair filtering and (B) filtered sample without occluding hairs

4.2.3 | Pixel-wise transformations

To achieve a visible variation in the intensity distribution of lesions, pixel-wise transformations were applied. The transformations involved are the CLAHE transform and the hue-saturation-value (HSV) transform. The visual improvement can be seen in Figure 6.

The overall combination of hair filtering and pixel-wise transformations provides the balanced benefits of both techniques. The resultant image can be seen in Figure 7.

4.3 | Proposed network

The two main stages in the proposed network are the segmentation branch and the classification branch. The segmentation network aims to extract the lesion-affected regions of the image, excluding any skin or occluding hairs. So, the segmentation was performed with basic binary classes in this case: background and foreground. After extraction, the segmented images are sent as input to the classification network. This network aims to precisely classify all seven classes of skin lesions.

4.3.1 | Stage 1—segmentation network

The implemented segmentation network was inspired and constructed to combine the basis of residual networks and atrous convolutions. The proposed model is made up of 26 layers, with over 15 million parameters.

In a residual network, each layer feeds into the next layer and directly into the layers further away using skip connections. Through this, the vanishing gradient problem and accuracy saturation problem is avoided by using residual blocks while training. Atrous convolutions increase the
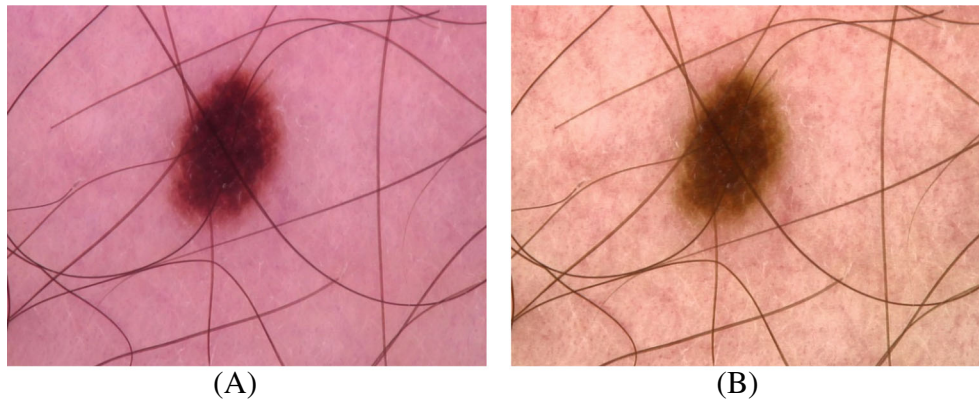
FIGURE 6 (A) Sample before pixel-wise transformations and (B) transformed sample with improved discrimination
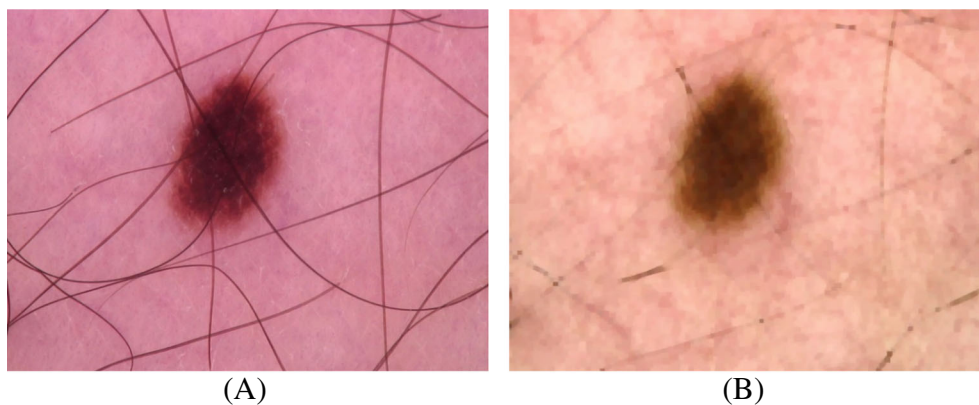
FIGURE 7 (A) Sample before hair filtering + pixel-wise transformations and (B) filtered + transformed sample
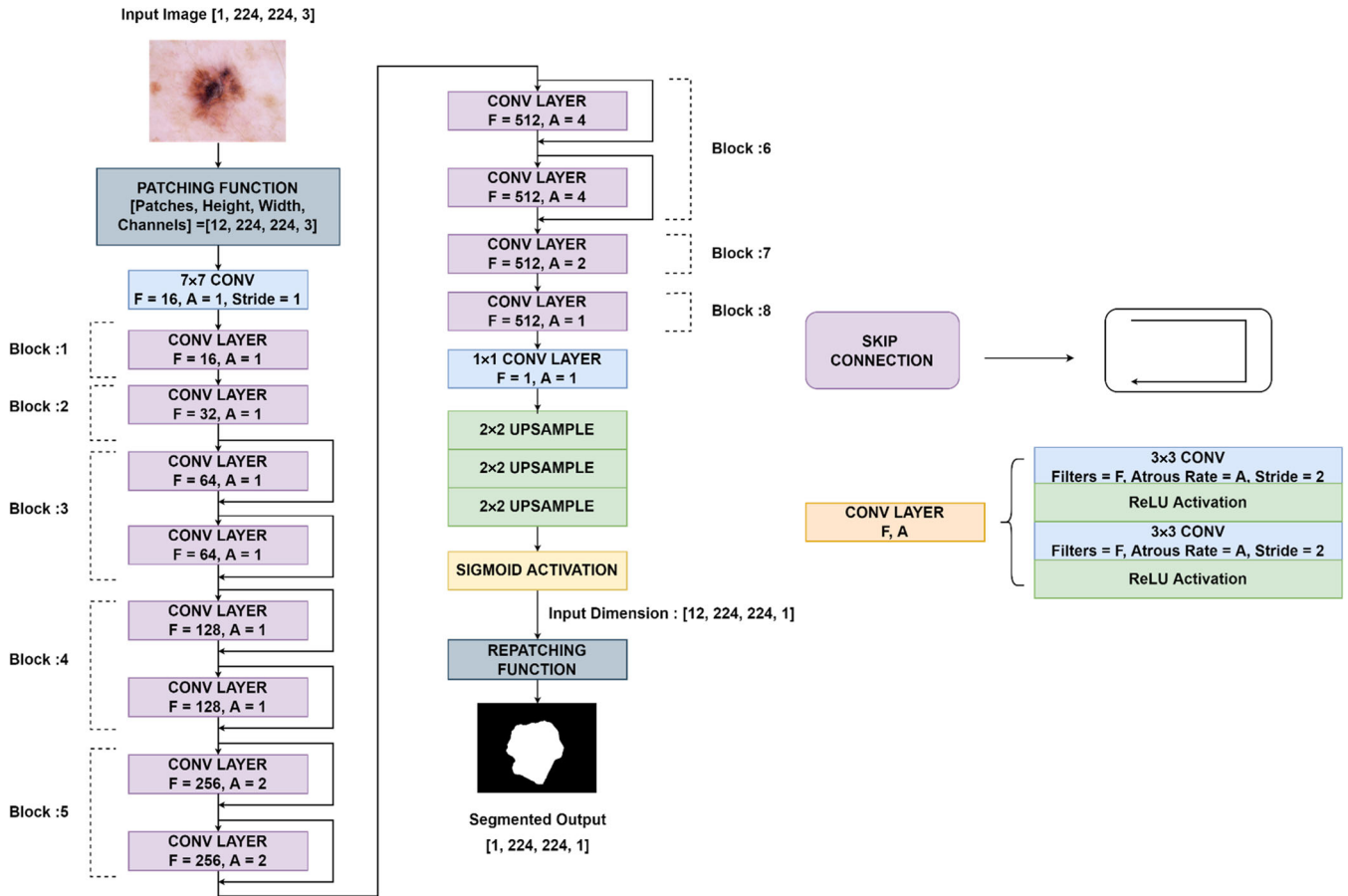
FIGURE 8 Architecture diagram of the proposed segmentation network with sub-block architecture.

overall receptive field of the filters. This allows for more information to be extracted without losing out on resolution. A global context is maintained and retains any significant information extracted.[33]

In segmentation tasks, the extracted features are generally upsampled after downsampling due to convolutions. The features are upsampled to the original size of the image to report the final segmented mask. However, if the downsampled image is of low resolution, the segmentation of the object will not be of good quality. In the original ResNet architecture, the final feature maps are of low dimensionality. Lower dimensionality is concurrent with low-quality feature extraction, which in turn affects the quality of the segmentation. Thus, atrous convolutions are a replacement for subsampling and maintain the size of the receptive field, without losing the context. Hence in the proposed segmentation network using ResNet, atrous convolutions are included in the place of normal convolution operations. This will allow the network to generate high-quality feature maps for precise segmentation.

However, the advantages of residual skip connections are maintained. They provide a way to retain the spatial resolutions through the end. The proposed model draws some inspiration from the combination of atrous convolutions and a ResNet-based architecture presented in Reference 33. The overall architecture of the proposed segmentation network is represented in Figure 8. It is designed using the basic blocks illustrated in the figure as well. This network involves deep layers to perform the initial feature extraction, leading to downsampling. After reaching a base dimension of $28 \times 28 \times 1$, the resultant feature map is upsampled and activated to obtain the segmented mask in 12 patches. Further, these patches are subjected to repatching, to obtain the final predicted mask. Throughout the model, the varying atrous rates in the intermediate stages of the network ensure that the key features are retained at specific layers.

The algorithm for the proposed segmentation network is presented as follows (Algorithm 1).

Algorithm 1. Segmentation neural network training

---

1. **Input**: D_t, training dataset; D_v, validation dataset; T stopping error threshold; *max_epochs*: upper limit on training epochs; γ, learning rate.
2. $W, b \leftarrow$ Uniform Xavier initialization. W, b are parameters and bias terms of the segmentation neural network.
3. $L_v \leftarrow \infty$, is the Stochastic gradient loss on the validation dataset.

4. *epoch* ← 0, tracks the epoch number.

5. **while** *epoch* < *max_epochs* and $L_v > T$ do

6.   Sample image-mask pair (I_t, M_t) from the training set D_t.

7. $P \leftarrow PatchSplit(I_t)$ divides the image into m patches.

8.   **for** p in P do

9. $x \leftarrow CONV\_series(p)$

10. $x \leftarrow UPSAMPLE\_series(x)$

11. $x \leftarrow Sigmoid(x)$

12. $M_p \leftarrow$ Collect and merge transformed patches

13. $L \leftarrow 1 - \frac{2|M_p \cap M_t|}{|M_p| + |M_t|}$

14. $W \leftarrow W - \gamma \frac{dL}{dW}$

15. $b \leftarrow b - \gamma \frac{dL}{db}$

16. $epoch \leftarrow epoch + 1$

17.   Compute dice loss L_v on the validation set D_v

Return W, b.

4.3.2 | Stage 2—classification network

The proposed classification network is similar to the employed segmentation network in terms of its basic architecture. However, the number of layers and atrous rates vary as per the classification requirements. The proposed classification network is made up of 42 layers and over 41 million parameters. A deeper model was chosen for the benefit of precise feature extraction. The proposed model was characteristically beneficial in spatial feature extraction, and hence was the preferred choice for the classification network as well (in comparison to default convolution networks).

Furthermore, the classification was modified in terms of additional convolutional and dense layers to suit the task at hand. Batch normalization and dropout layers were included to promote regularization as seen in Figure 9. The segmented images are categorized, preprocessed and balanced. The initial feature extraction leads to an extent of downsampling of feature maps. However, the varied atrous rates ensure that the global context is maintained. This is a key feature in deep architectures like the proposed classification network. The additions of normalizations, pooling, and fully connected layers simply allow the network to satisfy the classification task at hand.

The algorithm for the proposed classification network is presented as follows (Algorithm 2).

Algorithm 2. Classification neural network training

1. **Input**: D_t, training dataset; D_v, validation dataset; T stopping error threshold; *max_epochs*: upper limit on training epochs; γ, learning rate; K classes.

2. $W, b \leftarrow$ Uniform Xavier initialization. W, b are parameters and bias terms of the classification neural network.

3. $L_v \leftarrow \infty$, is the Stochastic gradient loss on the validation dataset.

4. *epoch* ← 0, tracks the epoch number.

5. **while** *epoch* < *max_epochs* and $L_v > T$ do

6.   Sample image, mask, one-hot class label triplet (I_t, M_t, y) from the training set D_t.

7. $x \leftarrow CONV\_net(I_t \otimes M_t)$

8. $\hat{y} \leftarrow Softmax(x)$

9. $L \leftarrow \sum_{i=1}^{K} - y_i \log(\hat{y}_i)$

10. $W \leftarrow W - \gamma \frac{dL}{dW}$

11. $b \leftarrow b - \gamma \frac{dL}{db}$

12.   Compute categorical cross-entropy loss L_v on the validation set D_v

13. $epoch \leftarrow epoch + 1$
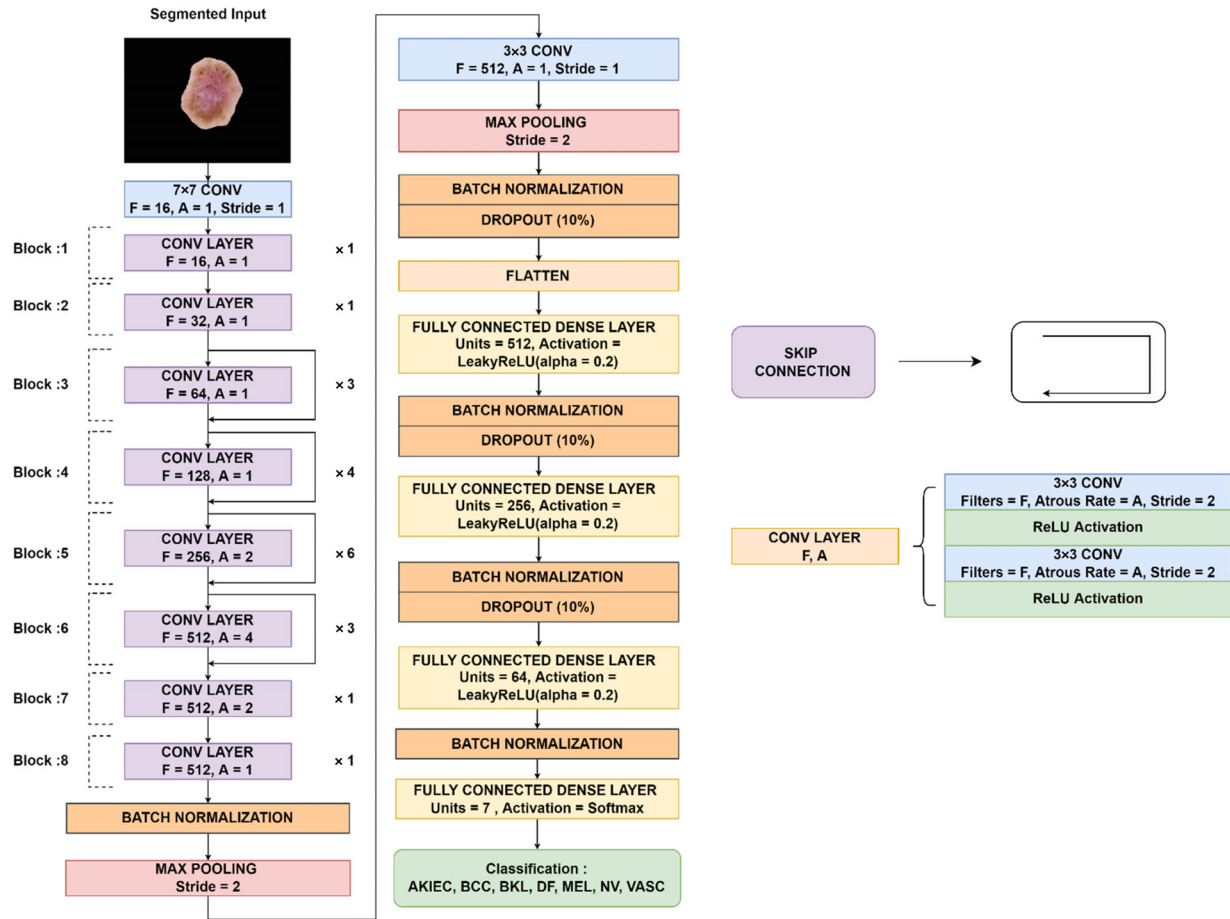
14. **Return** W, b

FIGURE 9 Architecture diagram of the proposed classification network with sub-block architecture.

5 | RESULTS AND DISCUSSION

5.1 | Experimental setup

The experiments conducted in this study were performed in a Python 3 environment on the Keras and Tensorflow frameworks. The computational resources required for the training of the proposed model involved a Tesla K80 GPU. These networks employed an ADAM optimizer and a learning rate of 0.0001 with the sparse categorical cross-entropy loss. The dataset employed was split into training and testing sets on an 80:20 split ratio.

5.2 | Evaluation metrics

Classification accuracy (ACC), precision (PRE), recall (REC), and F1 score are the measures utilized to assess the performance of our proposed classification models. From Equations (3)–(6), the equations for the aforementioned evaluation metrics are provided below.

$$ACC = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \tag{3}$$

$$PRE = \frac{T_P}{T_P + F_P}, \tag{4}$$

$$REC = \frac{T_P}{T_P + F_N}, \tag{5}$$

$$F1\ score = 2 \times PRE \times \frac{REC}{REC + PRE}, \tag{6}$$

where T_P stands for true positive, T_N stands for true negative, F_P stands for false positive, and F_N stands for false negative.

Additionally, we used intersection over union (IOU) and dice scores (DSC) to evaluate the results of our segmentation models. DSC measures the overlapping volume between two segmentations, as seen in Equation (7).

$$DSC = \frac{2\,|\,A \cap B\,|}{|\,A\,| \pm |\,B\,|},$$ (7)

where A and B denote the voxel sets for segmentation and ground truth, respectively.

IOU or Jaccard index is determined as the area of overlap between the prediction and actual ground truth, divided by the area of union between the prediction and ground truth. Equation (8) illustrates this.

$$JI = \frac{I_G \cap I_R}{I_G \cup I_R},$$ (8)

where I_G = ground truth image and I_R = predicted segmentation map.

5.3 | Ablation studies

This section details the ablation studies conducted to evaluate the contributions of the proposed work. Through this study, the significance and value of each component are analyzed, for a better understanding of the model's working.

5.3.1 | Analysis of preprocessing and augmentation

In the first section of the ablation studies, the performance of the preprocessing and the augmentation techniques proposed was analyzed. Each technique was included sequentially on a pretrained segmentation architecture. The U Net architecture was used for the ablation experiments. These results can be seen in Table 3.

From Table 3, it could be inferred that the performance of the segmentation using U-Net was better when the preprocessed dataset is considered for training. Also, to further improve the segmentation performance, atrous and residual layers were included in the backbone network and the related observations were presented in the next subsection.

5.3.2 | Analysis of the proposed segmentation network

The proposed segmentation network aimed to leverage the advantages of atrous residual networks to report precise segmentation performance. The observations from the ablation study can be seen in Table 4. Each technique was sequentially incorporated into a "backbone network" and evaluated for its contribution. This backbone network represents the architecture from Figure 9 where all skip connections are removed, and all atrous rates are equal to one. The proposed network was trained for 20 epochs. Figure 10 illustrates the convergence of the dice score for training and validation sets.

From Table 4, it could be inferred that the proposed segmentation network was able to segment the skin lesions with a dice score of 91.33%. The value addition brought by the introduction of residual and atrous convolution layers could be observed from the above observations. These layers were able to extract precise feature patterns from the input samples and exhibit better segmentation performance when compared to the backbone network. The illustration comparing the input image, ground truth mask, predicted mask, and predicted segmented output can be found in Table 5. The predicted masks were almost closer to the ground truth mask, exhibiting good segmentation performance.

TABLE 3 Observations of ablation results for the proposed preprocessing and augmentation.

S. no	Model used	Dice score (in %)	IoU score (in %)
1.	U Net—trained on native dataset	49.32	41.81
2.	U Net—trained on hair filtered dataset	55.35	47.68
3.	U Net—trained on augmented dataset	39.30	33.12
4.	U Net—trained on hair filtered + augmented dataset	55.24	49.76

TABLE 4 Observations of ablation results for the proposed segmentation network.

S. no	Model used	Dice score (in %)	IoU score (in %)
1.	Backbone network	87.63	80.23
2.	Backbone network including residual layers	88.44	80.91
3.	Backbone network including atrous layers	89.16	81.86
4.	Proposed network	91.33	84.69
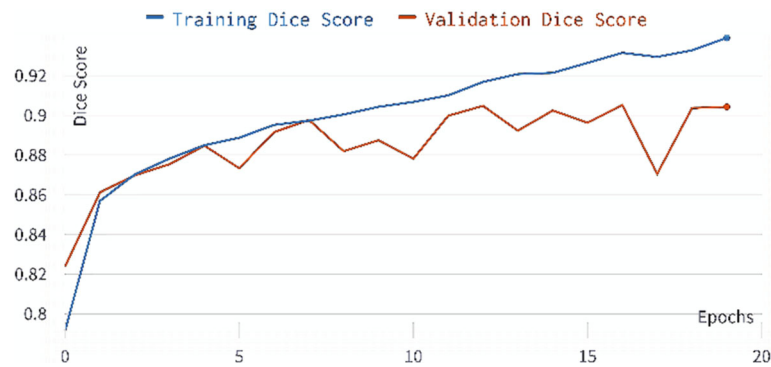
FIGURE 10 Observations of the training phase for the proposed segmentation network.

5.3.3 | Analysis of the proposed classification network

The proposed classification network aimed to leverage the advantages of atrous residual networks as well, with modifications to suit a classification task. Each technique was sequentially incorporated to a "backbone network" and evaluated for its contribution. This backbone network represents the architecture from Figure 9 where all skip connections are removed, and all atrous rates are equal to one. The proposed classification network achieved benchmark inter-class classification accuracy. As seen in the confusion matrix in Figure 11., the overall distribution of true classifications is higher, with lower false classifications.

The observations from the ablation study are presented in Table 6. It provides an overview of the incremental evaluation of the proposed system in different stages, starting from the back bone network. From Table 6, it could be observed that the performance of the backbone network was highly improved due to the integration of residual layers and refinement of the proposed network with separate segmentation and classification heads. The inter-class performance of the proposed network, and its ablation study can be seen in Table 7. It could be observed that the proposed network performs reasonably well for all the classes. Also, there is a room for improvement for "BKL" and "MEL" classes.

The observed loss and accuracy graphs for training and validation sets can be seen in Figure 12. As illustrated the validation accuracy has shown improvement over the training process. Both graphs have reached optimal convergence over 25 epochs. The precision-recall and ROC curves are illustrated in Figure 13.
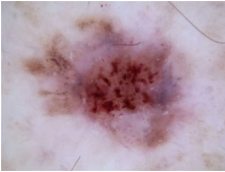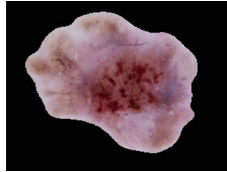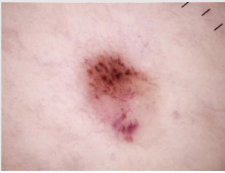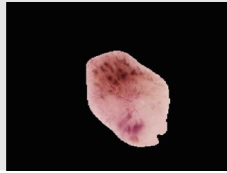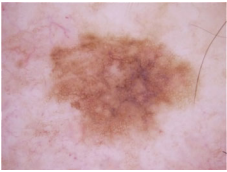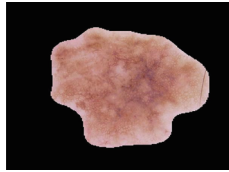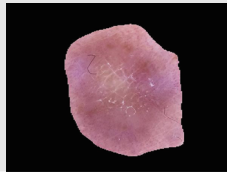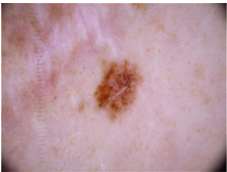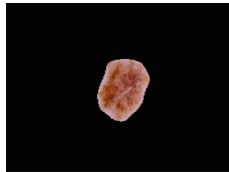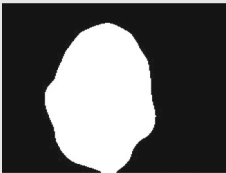
5.4 | Performance analysis

In the following section, the two proposed stages are individually compared to popular pretrained models. This comparison is performed to provide a benchmark for the overall performance of the proposed detection approach.

5.4.1 | Comparison of the proposed segmentation network with pretrained models

The implemented network was compared with popular segmentation architectures on the same dataset. This provides a comparison of our performance to the benchmark of pretrained models. The performance comparison of the proposed segmentation model to pretrained arch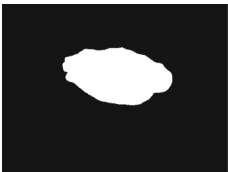itectures can be seen in Table 8. It could be inferred that the proposed segmentation network was able to outperform the existing networks employed for segmentation. This is due to the incorporation of residual and atrous convolution layers in the network for extracting precise feature patterns of each class.

TABLE 5 Ground truth versus predicted segmented samples of each class of the HAM10000 dataset after passing through stage 1.

S. no	Class	Input image	Ground truth mask	Predicted mask	Segmented output
1.	AKIEC				
2.	BCC				
3.	BKL				
4.	DF				
5.	MEL				
6.	NV				
7.	VASC				
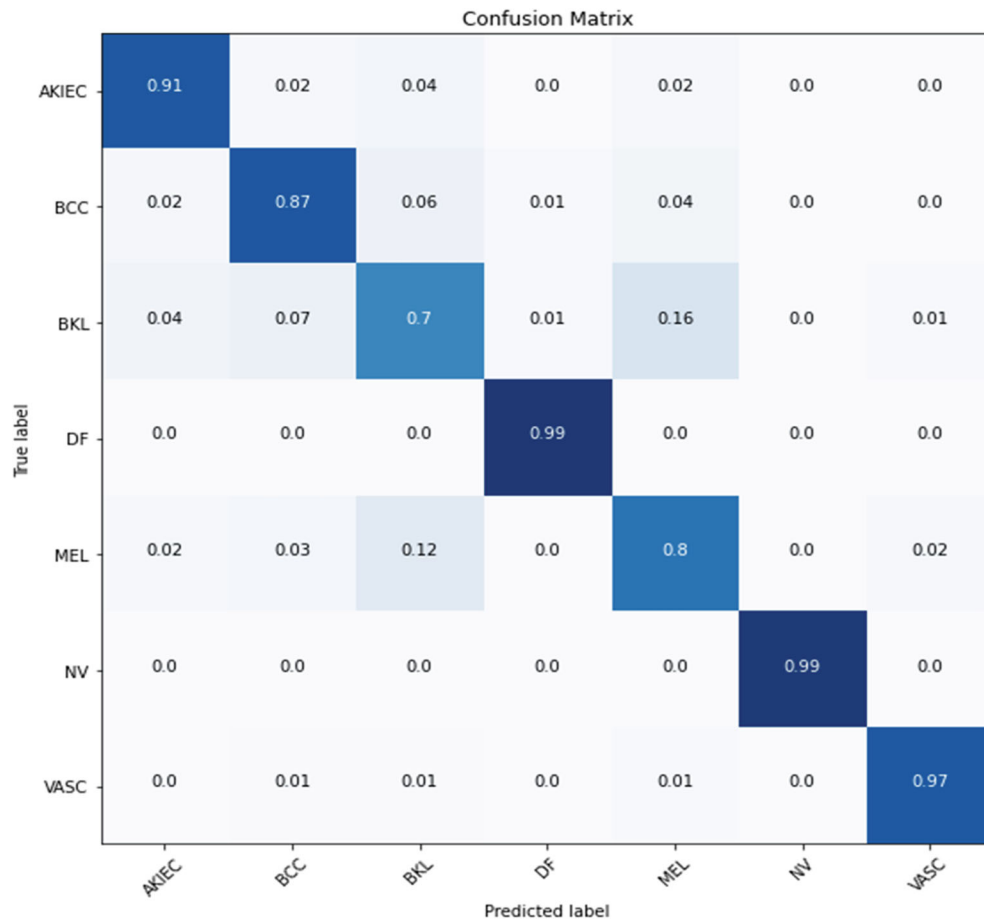
FIGURE 11 Confusion matrix of the classification model.

TABLE 6 Summary of the ablation studies made.

S. no	Model used	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC score (%)
1.	Backbone network	59.80	62.14	59.75	59.02	89.94
2.	Backbone network including residual layers	74.17	77.10	74	74.98	95.14
3.	Proposed model	89.27	89.06	89.18	89.09	98.48

TABLE 7 Observations of the class-wise metrics for the proposed classification network.

Class/metric	Precision (%)	Recall (%)	F1-score (%)
AKIEC	90.54	91.5	91.02
BCC	87.93	86.74	87.33
BKL	75.71	70	72.74
DF	97.71	98.93	98.32
MEL	76.21	80.41	78.25
NV	99.18	99.36	99.27
VASC	96.15	97.33	96.73

FIGURE 12 Observations of training versus validation sets for classification (A) loss and (B) accuracy.

FIGURE 13 (A) Precision-recall and (B) ROC curve for the seven classes of the HAM10000 dataset

TABLE 8 Performance comparison of proposed segmentation model with pretrained models.

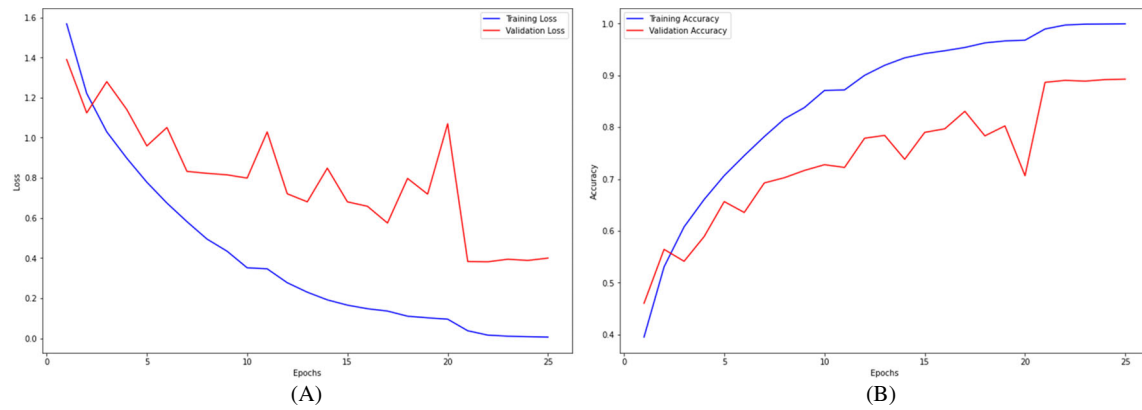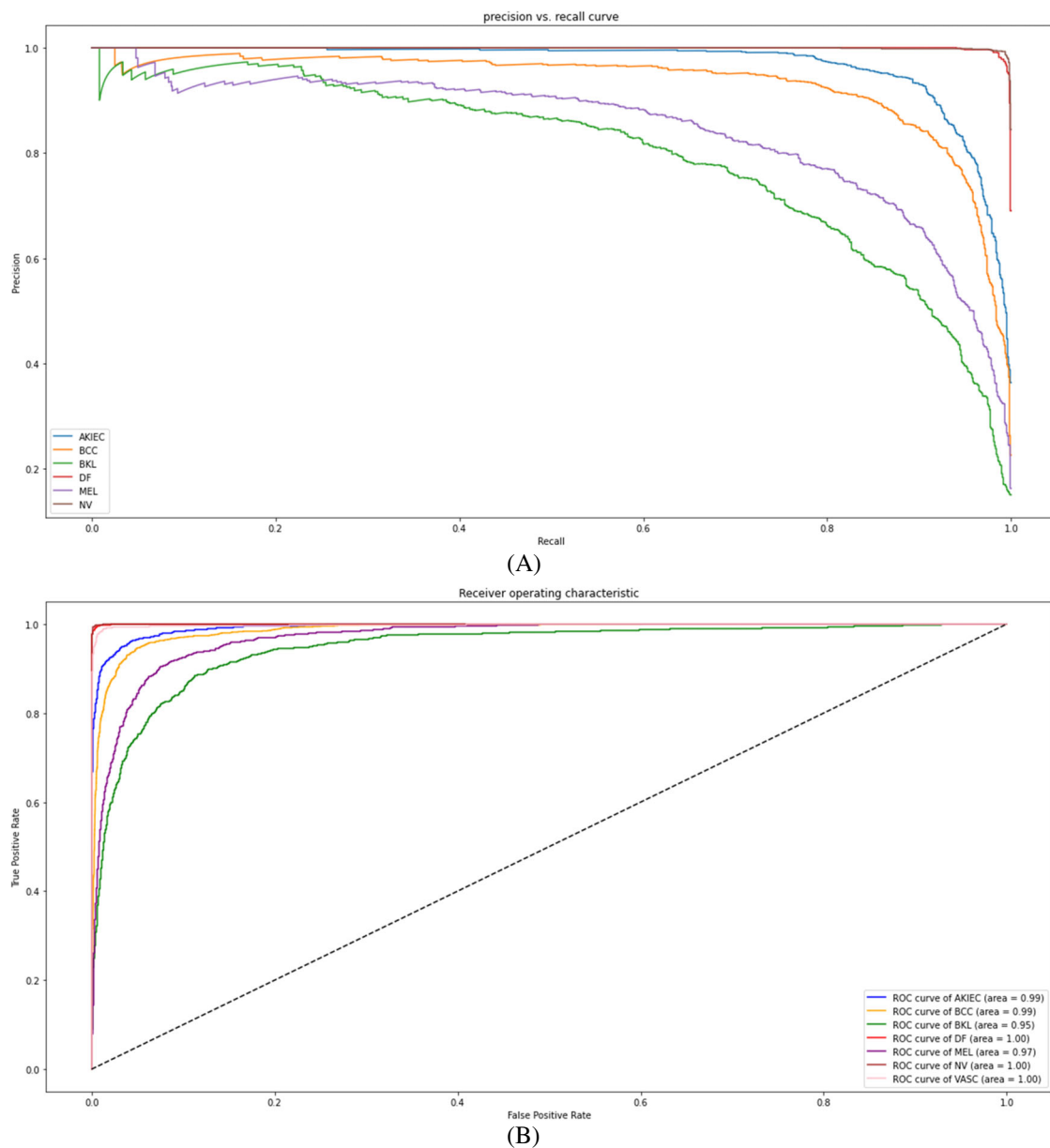| S. no	Model used	Dice score (in %)	IoU score (in %)
1.	U Net	49.32	41.81
2.	PSPNet	40.73	30.36
3.	LinkNet	55.51	47.23
4.	Proposed model	89.69	82.11

TABLE 9 Performance comparison of proposed classification model and pretrained models.

S. no	Model used	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1.	Mobilenet V2	44.5	24.87	56.86	16.86
2.	Xception	74.17	78.29	74.17	73.9
3.	DenseNet	75.88	78.3	75.8	75.9
5	Inception	77.27	79.11	77.3	76.7
5.	Proposed model	89.27	89.06	89.18	89.09

TABLE 10 Performance comparison of the proposed model with existing works.

S. no	Model used	Number of classes	Accuracy (%)
1.	Karthik et al.[15]	4	84.7
2.	Połap et al.[18]	2	89.12
3.	Khan et al.[19]	2	89.12
4.	Khan et al.[20]	3	94.3
5.	Goyal et al.[26]	2	93.8
6.	Agrahari et al.[10]	7	80.81
7.	Waweru et al.[16]	7	86
8.	Cengil et al.[22]	7	77.80
9.	Khan et al.[25]	7	86.1
10.	Proposed model	7	89.27

5.4.2 | Comparison of the proposed classification network with pretrained models

The implemented network was compared with popular classification architectures on the same dataset. This provides a comparison of our performance to the benchmark of pretrained models. The performance comparison of the proposed classification model to pretrained architectures can be seen in Table 9. It could be inferred that the proposed classification network was able to outperform the existing state-of-the-art pretrained networks.

5.4.3 | Comparison of the proposed network with existing works

The implemented network was compared with the studies discussed in Section 2. This provides a comparison of our performance to the related works. The performance comparison of the proposed classification model against the existing works can be seen in Table 10. The initial records illustrate selected works[15,18–20,26] that have employed a lower range of skin disease types in their classification tasks. While these works may report higher values of performance, they lack the inter-class balance. Furthermore, the performance of works that have employed seven classes are compared with the proposed model as well.[10,16,22,25] Though the performance of these works were better, it often overlooks inter-class performance. The proposed network was able to obtain optimal results with an accuracy of 89.27% without compensating for the inter-class performance.

6 | CONCLUSION

Skin lesions are known to increase the risk of developing skin diseases. Owing to its perception as a low-risk disease, lesions are often overlooked. This attributes to a delay in diagnosis, where the undiagnosed skin lesions have the possibility of manifesting as skin cancer. The more the delay, the higher the risk to the patient's life. Early detection of skin lesions and its severity is essential for starting necessary treatment. Existing methods predominantly involve physical intervention by dermatologists. These techniques can be time-consuming, and carries a reasonable level of risk. The presented research leverages deep learning algorithms to segment lesion-affected regions, and classify them into seven different categories of cancerous and noncancerous lesions.

While many works employ deep learning for skin disease detection, there are a few research gaps to be addressed. These include improving discrimination, augmentation, and precise segmentation of lesion features. The proposed model is trained on the open-source HAM10000 dataset with seven classes of skin lesions. The proposed approach presents two stages, dedicated to address the gaps discussed. This approach involves a novel combination of preprocessing, an atrous residual segmentation network, and a robust classification network. By performing patching as the first step in preprocessing, the segmentation is focused on the subtle morphological variations of the lesion. To further bolster the detection performance of residual learning, atrous convolutions maintain the contextual features extracted. Also, the use of preprocessing techniques plays a role in improving the discriminative robustness. The proposed method achieved precise segmentation, acting as a basis for improved detection performance. The proposed classification model achieved an accuracy, precision, recall, and F1-score of 89.27%, 89.06%, 89.18%, and 89.09%, respectively. The proposed model has outperformed state-of-the-art pretrained architectures tested on the same seven-class dataset. Future work for this research would involve incorporating other features like patient metadata, and region-based analysis to extend the scope of real-world automated skin lesion detection.

CONFLICT OF INTEREST STATEMENT

The authors declare no known potential competing interests for financial interests or the research, authorship, and publication of this article.

DATA AVAILABILITY STATEMENT

The HAM10000 dataset is available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T.

ORCID

Karthik Ramamurthy https://orcid.org/0000-0002-5250-4337
Gugan S. Kathiresan https://orcid.org/0000-0002-1277-3145

REFERENCES

1. Labani S, Asthana S, Rathore K, Sardana K. Incidence of melanoma and nonmelanoma skin cancers in Indian and the global regions. *J Cancer Res Ther*. 2021;17(4):906-911.
2. Preda-Naumescu A, Weir S, Sowell JW. 33632 educational interventions may improve sun protection and skin cancer awareness in outdoor occupational workers. *J Am Acad Dermatol*. 2022;87(3):AB48. doi:10.1016/j.jaad.2022.06.226
3. Wu YP, Parsons B, Jo Y, et al. Outdoor activities and sunburn among urban and rural families in a Western region of the US: implications for skin cancer prevention. *Prev Med Rep*. 2022;29:101914. doi:10.1016/j.pmedr.2022.101914
4. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. *Int J Cancer*. 2021;149:778-789.
5. Cancer Facts and Figures. American Cancer Society. Accessed January 19, 2022 2022 https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2022/2022-cancer-facts-and-figures.pdf
6. https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer#. Accessed February 1, 2022.
7. Centers for Disease Control and Prevention. Facts and statistics about skin cancer. 2006.
8. https://www.aad.org/public/diseases/skin-cancer/find/know-how#:~:text=Skin%20cancer%20diagnosis%20always%20requires%20a%20skin%20biopsy&text=The%20procedure%20that%20your%20dermatologist,way%20to%20know%20for%20sure. Accessed February 1, 2022.
9. Anand V, Gupta S, Koundal D. Detection and classification of skin disease using modified Mobilenet architecture. *ECS Trans*. 2021;107(1):5059.
10. Agrahari P, Agrawal A, Subhashini N. Skin cancer detection using deep learning. In: Sivasubramanian A, Shastry PN, Hong PC, eds. *Futuristic Communication and Network Technologies*. Springer; 2022:179-190.
11. Sae-Lim W, Wettayaprasit W, Aiyarak P. Convolutional neural networks using mobilenet for skin lesion classification. Proceedings of the 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE); 2019:242-247; IEEE.
12. Calderón C, Sanchez K, Castillo S, Arguello H. BILSK: a bilinear convolutional neural network approach for skin lesion classification. *Comput Methods Programs Biomed Update*. 2021;1:100036.
13. Khan MA, Javed MY, Sharif M, Saba T, Rehman A. Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS); 2019:1-7; IEEE.
14. Thomas SA. Combining image features and patient metadata to enhance transfer learning. Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021:2660-2663; IEEE.
15. Karthik R, Vaichole TS, Kulkarni SK, Yadav O, Khan F. Eff2Net: an efficient channel attention-based convolutional neural network for skin disease classification. *Biomed Signal Process Control*. 2022;73:103406.

16. Waweru AK, Ahmed K, Miao Y, Kawan P. Deep learning in skin lesion analysis towards cancer detection. Proceedings of the 2020 24th International Conference Information Visualisation (IV); 2020:740-745; IEEE.

17. Huo Y. Full-stack application of skin cancer diagnosis based on CNN model. Proceedings of the 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI); 2021:754-758; IEEE.

18. Połap D. Fuzzy consensus with federated learning method in medical systems. *IEEE Access*. 2021;9:150383-150392.

19. Khan MA, Muhammad K, Sharif M, Akram T, Kadry S. Intelligent fusion-assisted skin lesion localization and classification for smart healthcare. *Neural Comput Appl*. 2021;1-16.

20. Mahbod A, Tschandl P, Langs G, Ecker R, Ellinger I. The effects of skin lesion segmentation on the performance of dermatoscopic image classification. *Comput Methods Programs Biomed*. 2020;197:105725.

21. Khan MA, Muhammad K, Sharif M, Akram T, de Albuquerque VHC. Multi-class skin lesion detection and classification via Teledermatology. *IEEE J Biomed Health Inform*. 2021;25:4267-4275.

22. Cengil E, Çinar A, Yildirim M. Hybrid convolutional neural network architectures for skin cancer classification. *Avrupa Bilim Ve Teknoloji Dergisi*. 2021;28:694-701.

23. Attique Khan M, Sharif M, Akram T, Kadry S, Hsu CH. A two-stream deep neural network-based intelligent system for complex skin cancer types classification. *Int J Intell Syst*. 2021;37:10621-10649.

24. Khan MA, Sharif M, Akram T, Damaševičius R, Maskeliūnas R. Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. *Diagnostics*. 2021;11(5):811.

25. Khan MA, Zhang YD, Sharif M, Akram T. Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. *Comput Electr Eng*. 2021;90:106956.

26. Goyal M, Oakley A, Bansal P, Dancey D, Yap MH. Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*. 2019;8:4171-4181.

27. Sarker MMK, Rashwan HA, Akram F, et al. SLSNet: skin lesion segmentation using a lightweight generative adversarial network. *Expert Syst Appl*. 2021;183:115433.

28. Singh VK, Abdel-Nasser M, Rashwan HA, et al. FCA-net: adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access*. 2019;7:130552-130565.

29. Yang CH, Ren JH, Huang HC, Chuang LY, Chang PY. Deep hybrid convolutional neural network for segmentation of melanoma skin lesion. *Comput Intell Neurosci*. 2021;2021:9409508.

30. Ünver HM, Ayan E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics*. 2019;9(3):72.

31. Anand V, Gupta S, Koundal D, Mahajan S, Kant Pandit A, Zaguia A. Deep learning based automated diagnosis of skin diseases using Dermoscopy. *Computers, Materials & Continua*. Vol 71. Tech Science Press; 2022:3145-3160. doi:10.32604/cmc.2022.022788

32. Anand V, Gupta S, Koundal D, Nayak SR, Barsocchi P, Bhoi AK. Modified U-NET architecture for segmentation of skin lesion. *Sensors*. 2022;22(3):867. doi:10.3390/s22030867

33. Cui X, Zheng K, Gao L, Zhang B, Yang D, Ren J. Multiscale spatial-spectral convolutional network with image-based framework for hyperspectral imagery classification. *Remote Sens*. 2019;11(19):2220.