International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

GAUSSIAN DISTRIBUTION-BASED SPATIAL ATTENTION MAP ESTIMATION WITH ITERATIVE CHANNEL SELECTION FOR THE DETECTION OF SKIN LESIONS

Kavitha B[1] and Dr. Kusuma Kumari B M[2]

[1]Research Scholar and [2]Assistant Professor, Department of Studies and Research in Computer Applications,Tumkur University, Tumakuru, Karnataka

[1]kavithab@tumkuruniversity.ac.in and [2]kusuma_bm@tumkuruniversity.ac.in

ABSTRACT

This paper proposes a skin lesion region segmentation and classification approach that uses the U-Net tapped descriptors obtained from the different levels of the encoder. The architecture uses the U-Net structure as a segmentation network for detecting the lesion region, while blocks such as channel selection, Gaussian distribution-based spatial attention map, channel elimination, and channel duplication are used in the classification network. The traditional spatial attention map is utilized to obtain the dominant region of the image to extract the descriptors. However, the spatial attention (SA) is learned during the training process, which increases the complexity of the training process. To reduce such complexity, a Gaussian Distribution-Based Spatial Attention (GDSA) is proposed to compute the attention map from the input feature map with the use of a Gaussian distribution. The proposed spatial attention map is a non-learnable parameter that reduces the complexity in training. The paper also proposes an Iterative Channel Selection (ICS) process that selects the dominant feature channels while eliminating the redundant channels. The evaluation of the suggested (GDSA-U-Net) skin lesion classification approach was evaluated using the measures, namely accuracy, Matthews Correlation Coefficient(MCC), precision, F1-score, Kappa coefficient, and specificity, with the datasets, namely HAM-10000, and ISIC-2019. The suggested GDSA-U-Net structure yields a classification accuracy of 93.71% and 94.85% when evaluated using the HAM-10000 and ISIC-2019 datasets. In the case of the ISIC-2019 dataset, the suggested GDSA-U-Net yields an MCC, precision, and Kappa coefficient of 93.08%, 93.99%, and 94.14%, respectively.

Keywords: Spatial attention, U-Net, Feature channels, Skin lesion, Gaussian distribution

1. INTRODUCTION

The abnormal and irregular growth of skin cells leads to skin cancer [1], which is a deadly disease. Due to the spreading nature of the cancer cells through the blood and lymphatic system, the healthy tissue in the skin or body is destroyed. The death rate due to cancer has significantly increased over the past few years, and it is considered a rapidly growing cancer across the globe [2]. A survey by the International Agency for Research on Cancer (IARC) reports that in the year 2018, around 18 million new skin cancer cases were reported across the globe, while the deaths due to skin cancer were 9 million across the globe for the same year. The patient survival rate increases as skin cancer is diagnosed and treated at its early stage. For diagnosing skin cancer, dermatologists rely on clinical skin biopsy or dermoscopic pictures [3]. Diagnosing skin cancer using the dermoscopic pictures by the visual inspection of a dermatologist is a time-consuming procedure and is frequently error-prone. Recently, due to the emergence of artificial intelligence-based diagnostic systems, such

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

as the deep learning approach, has highly minimized the intra-class variation and inter-class similarity. Few skin cancer types require surgical removal since it is considered risky to the patients.

Different deep descriptors are extracted from the region of interest using different deep learning architectures, which are used in several classification applications [4, 5]. To minimize the effect of intra-class variation, inter-class similarity, and insufficient training images in Deep-Convolutional Neural Network (D-CNN), the authors Zhang et al. [6] introduced an attention residual learning (ARL) approach. This approach uses several ARL blocks followed by a pooling and classification process. To improve the discriminative power, attention learning and residual learning are used in each ARL layer. The attention process constructs an attention map required for low layers from the higher-layer feature map. Different convolutional neural network (CNN) structures are utilized to construct an ensemble CNN [7], where weighted outputs estimated from different CNNs are used to compute the classification result. This approach also uses different fusion approaches to aggregate the descriptors. The authors report that the use of an ensemble CNN significantly reduces the requirement for training data. To enhance the number of training images, the Generative Adversarial Network (GAN) structure [8] is used for augmentation. To generate skin images, the generator and discriminator control the noise and style. A transfer learning classifier is used to train the skin images generated by the GAN network. This approach provides an accuracy, precision, and specificity of 83.1%, 96.6%, and 74.3% when evaluated using the ISIC-2018 dataset.

Two networks, namely inter-network and intra-network, are used to construct an ensemble CNN [9] which contains different levels of descriptors extracted by different sets of pre-trained CNN structures. The extracted descriptors from different networks a((re trained and classified by Support Vector Machine (SVM), while the final prediction probability is computed from the average of the prediction probabilities computed on each network. This approach yields an AUC(Area Under Curve) of 0.873 and 0.955 for the melanoma type and Seborrheic keratosis skin lesions, respectively. Lucas et al. [10] used a fully convolutional network and a U-Net segmentation process from which deep and handcrafted descriptors are extracted. The use of a random forest classifier with a border-interior descriptor yields a maximum performance. Hasan et al. [11] used intensity and geometry-based augmentation in training a hybrid CNN that uses 3 unique descriptor extraction blocks formed with transfer learning models. The descriptors from these blocks are combined to obtain a depth descriptor map that better represents the lesion region. This hybrid CNN was named Dermo Expert, which yields an AUC of 0.97, 0.95, and 0.96 when evaluated using the ISIC-2018, ISIC-2017, and ISIC-2016 datasets. Sathvika et al. [12] used SVM and AlexNet to derive a pipelined structure that extracts the bi-sectional texture descriptors to differentiate the skin lesion types.

To minimize over-fitting, a regularized Drop Block and Dropout module was used in Deep CNN [13]. This approach also uses a multi-weighted loss function to compensate non-uniform distribution of training data. Though the complexity of this approach is less, the performance is limited and can be improved. To categorize 7 types of skin lesions, Razia et al. [14] proposed the S-MobileNet architecture, which uses segmentation-based fractal texture analysis and a Gaussian filtering process for segmenting the lesion region. The approach uses an intermediate layer for the compression process that reduces the algorithm's complexity. The authors also report that replacing ReLU activation with Mish activation improves the classification performance. The CNN structure

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[15] was enhanced with the use of an attention branch. Similar to other spatial attention processes, the use of the attention branch helps to focus on a certain lesion region, which improves the discriminative power. The model minimizes the class imbalance problem with the use of entropy-guided loss. This approach results in an AUC of 0.922 and 0.719 when evaluated using the ISIC-2017 and ISIC-2016 datasets.

The global descriptors extracted from the transformer network are combined with the local descriptor extracted by the deep CNN to derive the CR-Conformer [16]. This dual-branch CNN also applies the convolution in a rotating order to obtain enhanced descriptors. The advantage of this approach is that it utilizes a smaller number of parameters, around 11.17M. The malignant and benign skin lesions are categorized by Ramesh et al. [17], where the region of interest can be localized using a swarm intelligence process. This swarm intelligence approach uses the Grasshopper optimizer. From the segmented region, the SURF descriptors are extracted and used for lesion classification. This approach results in an average MCC of 97.04% in malignant, benign lesion classification. The descriptors extracted from several deep models are used by the binary bat selection approach [18]. This optimization computes the descriptors that are highly distinctive based on the entropy. For the extraction of descriptors, architectures such as NasNet mobile, DenseNet-201, and Inception-ResNet v2 are used. The use of the optimizer not only minimizes the redundant descriptors but also considers the significant descriptors. This approach yields an accuracy of 95.71% when evaluated using the ISIC-2017 dataset.

Multi-scale dilation attention [19] was proposed to derive multi-residual attention. Different contextual cues in the dermoscopic pictures can be collected by the dilation attention. This attention compensates for lesion structure and size changes. Along with the residual attention channel, the spatial attention is used to detect the asymmetric and irregularities. To obtain discriminative descriptors from different channels, the squeeze and excitation process is used. The use of residual attention provides a lesion classification accuracy of 88% and 90% when analyzed using the ISIC-2019 and HAM-10000 datasets. Spatial Aware Squeeze Excitation (SWSE) process is used in the vision transformer [20] that extracts the global descriptor with the use of the EfficientNet B0 model. The use of SWSE attention focuses on essential components of the dermoscopic picture to differentiate the lesion type. This approach results in a recall and accuracy of 95% and 94.64%, respectively, when evaluated using the HAM-10000 database. An optimized relation-aware Recurrent Neural Network (RNN) [21] was proposed by Lakshmi et al., which uses a Cubature Kalman filter as pre-processing. The multi-view clustering process is then used to segment the lesion-affected region from which color, texture, and discriminative descriptors are extracted utilizing a squeezing transform. Further, the descriptors are optimized using black winged kite optimizer that reduces the complexity by estimating weights efficiently.

Most of the attention-based approaches that are discussed in the literature use a spatial or channel-based attention approach that computes and updates an attention map during the training process. This increases the complexity of the model during the training process due to the updation of the attention map during backpropagation (attention map values are learnable). To reduce the complexity in attention map computation, the paper proposes a non-learnable attention map that computes the map using the input feature map. An iterative channel selection approach is proposed

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

that selects the dominant descriptor channels while neglecting the redundant descriptor channels. The contribution of the work is as follows

(i) The work proposes a skin lesion segmentation and classification approach that was derived from the U-Net approach, which can differentiate the lesion types, namely Basal cell carcinoma, vascular lesion, actinic keratoses, Squamous cell carcinoma, dermatofibroma, melanoma, melanocytic nevi, and Benign keratosis.

(ii) The derived structure has two networks, namely the segmentation network and the classification network, while the U-Net structure is used in the segmentation network. The classification network uses a Gaussian distribution-based spatial attention that computes the attention map using the descriptor tapped from the U-Net.

(iii) Instead of using a channel attention that computes a map value for each channel, an Iterative Channel Selection (ICS) approach is used, which also eliminates a few channels and computes the 1D-channel weights. This channel selection process considers only the discriminative feature channels, eliminating other channels.

(iv) Finally, the evaluation of the suggested GDSA-U-Net approach was analyzed utilizing different evaluation scales with the use of HAM-10000 and ISIC-2019 datasets.

The following sections of the paper are constructed as follows. The working of the suggested GDSA-U-Net-based skin lesion segmentation and classification approach was presented in Section 2. The experimental analysis of the GDSA-U-Net structure was depicted in Section 3, while the conclusion of the work is presented in Section 4.

2. PROPOSED METHOD

The suggested GDSA-U-Net-based approach initially pre-processes the dermoscopic pictures to remove the hair artifacts and to enhance the skin image. The approach uses a hair artifact removal algorithm to remove the hair region present in the image. The GDSA-U-Net has two networks, namely the segmentation network and the classification network. The objective of the segmentation network is to segment the region of interest while eliminating other regions. Similarly, the classification network categorizes the segmented skin lesion region into one of the eight types, namely Basal cell carcinoma (BCC), vascular lesion (VAS), actinic keratoses (AKI), Squamous cell carcinoma (SCC), dermatofibroma (DEF), melanoma (MEL), melanocytic nevi (NV), and Benign keratosis (BKL). The skin lesion region segmentation algorithm uses a U-Net structure, while the lesion classification approach uses the U-Net tapped descriptors. The classification network has the modules, namely the channel selection (CS), Gaussian distribution-based spatial attention (GD-SA), Channel duplication (CD), Channel elimination (CE), weighted fusion, and a fully connected network as illustrated in Fig. 1.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

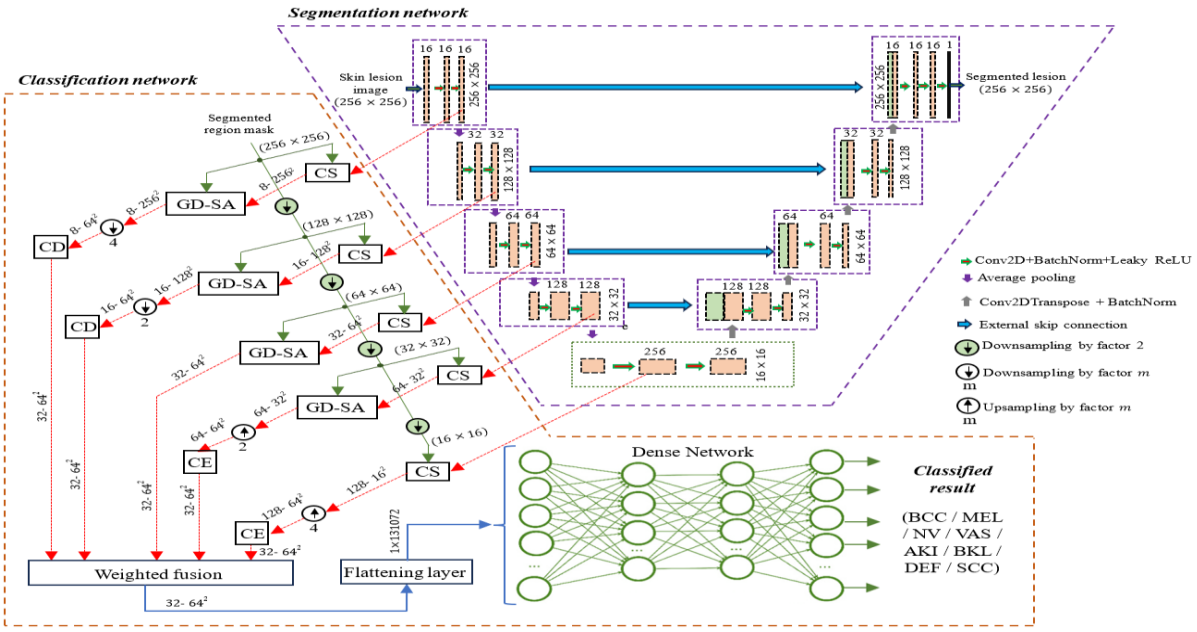


Fig. 1: Architecture of the GDSA-U-Net in detecting and classifying the skin lesions

2.1 Pre-Processing

Let $S_1(x, y)$ resembles the input skin image. Stages such as CLAHE(Contrast Limited Adaptive Histogram Equalization) enhancement [22] and morphological operation-based hair artifact removal [23] are used in the pre-processing stage. In the CLAHE image enhancement process, tiles are initially computed by fragmenting the image into non-overlapping blocks, where each block has a dimension of $\nabla \times \nabla$. Using a clip limit of ε, the normalized histogram computed on the tile λ_j is clipped using the relation,

$$\lambda_j^{(clip)} = \begin{cases} \varepsilon & if\ \varepsilon < \lambda_j \\ \lambda_j & if\ \ \varepsilon \geq \lambda_j \end{cases} \tag{1}$$

Here, j resembles the intensity present in an input image. The value of j lies between 0 to 255 for an 8-bit skin picture. The enhanced skin picture can be obtained using the relation $S_2(x, y) = G(S_1(x, y))$, where the function $G(.)$ can be computed using the relation,

$$G_j = \left\lfloor \frac{266\, c_j}{\nabla^2} \right\rfloor \tag{2}$$

where, c_j resembles the tile cumulative distribution. The presence of hairs over the lesion region or region of interest reduces the classification performance. Morphological operations involving dilation and erosion, and adaptive thresholding are utilized in the hair removal process [23]. Initially, the RGB picture is converted to grayscale, and the hair regions are detected using adaptive thresholding. The connected components are then detected using the morphological operations, from which the thickness and area of the segmented hair region are detected. The region that contains the hair is inpainted using a fast-marching approach to obtain the hair artifact-removed pictures $S_3(x, y)$. This preprocessed image $S_3(x, y)$ is used in segmenting and classifying the skin lesions.

2.2 Segmentation Network

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The skin lesion segmentation network uses a U-Net [24] structure that contains a bottleneck layer between the encoder and decoder sections. The encoder part contains a sequence of convolutional filters followed by a max-pooling layer. The number of channels in the convolutional filter increases as the feature propagates through the encoder part, however, the spatial dimension gets reduced. Usage of the features extracted by the encoder or contraction path of the U-net avoids the requirement of separate convolutional filters in the classification network. The expansion path mainly serves two purposes. Firstly, it combines the features of the decoder and encoder with the use of skip connections. Secondly, the reduced spatial dimension of the feature is recovered by the expansion path with the use of transposed convolutional filters that perform an upsampling operation. Let $M(x, y)$ resembles the segmented region mask detected by the U-Net segmentation network.

2.3 Classification Network

Instead of using a feature extraction network in the classification network, the suggested classification network uses the U-Net tapped descriptors. However, modules, namely channel selection, Gaussian distribution spatial attention (GD-SA), Channel duplication (CD), Channel elimination (CE), weighted fusion, and a fully connected network, are used in the classification network. Along with these modules, downsampling by 2 and downsampling by *m* modules are used to adjust the spatial dimension of the descriptor maps.

(i) Channel Selection

The channel selection process will select 50% of the channels that have discriminative features. An iterative channel selection (ICS) approach was proposed, which computes the dominant discriminative features. The selection was done using the energy parameter δ_j computed on each channel j. Let the input to the channel selection module have R channels. The channel selection module selects a discriminative $R/2$ channels. The channel selection was done based on the average gradient energy computed on the 4 neighbourhoods in the region of interest. The region of interest is the region that was enclosed by the U-net segmented region. At any point (x, y), the energy of the gradient that corresponds to the 4-neighbourhood is expressed as,

$$E(x,y) = \frac{1}{4}\left|\sum_{l\in\{-1,1\}}\sum_{m\in\{-1,1\}} D(x,y) - D(x-l, y-m)\right|^2 \qquad (3)$$

For a particular channel, the energy parameter for channel j is expressed as,

$$\delta_j = \widehat{M}(x,y)\frac{1}{z_1 \times z_2}\sum_{p=1}^{z_1}\sum_{q=1}^{z_2} E(p,q) \qquad (4)$$

Where $z_1 \times z_2$ is the size of the feature map $D(x, y)$ and $\widehat{M}(x, y)$ is the scaled version of the segmented mask image $M(x, y)$. Let δ_1, δ_2…..δ_R represents the energy parameter computed on R channels. The energy parameters are sorted from the lowest to the highest value. Let the sorted energy parameter be represented $\hat{\delta}_1$, $\hat{\delta}_2$…..$\hat{\delta}_R$. Here $\hat{\delta}_1$ resembles the least energy parameter and $\hat{\delta}_R$ is the highest energy parameter. The sorted energy parameter is then normalized between 0 and 1 such that the minimum value is 0 and the maximum value is 1. The normalized energy parameter for ith position can be represented as

$$\sigma_i = \frac{\hat{\delta}_i - \hat{\delta}_1}{\hat{\delta}_R - \hat{\delta}_1} \qquad (5)$$

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fig. 2 shows an example where the number of channels before channel selection is $R = 16$. The iterative channel selection (ICS) process will select $R/2 = 8$ channels. In this example, Row-1 resembles the normalized energy parameter values σ_i. Initially (in iteration 1), the channels that correspond to higher and lower energy parameters are selected, since it has higher discriminative energy parameter components. In iteration 2, the channel that has an energy parameter near to the average of the adjacent selected energy parameters is again selected. This process is repeated till the iteration satisfies the condition $N_{s,j} > R/2$. Here $N_{s,j}$ resembles the number of selected parameters till iteration j. In this example, the iteration reaches the stopping condition at $j = 4$ (iteration 4). Once the iteration stops, refining is performed so that the $N_{s,j} = R/2$. For performing the refining process, the difference between the adjacent energy parameters is computed, from which the energy parameter difference on the left and right sides is averaged. The channel that corresponds to the averaged adjacent energy parameter difference that has a higher $R/2$ discrimination (difference) is chosen as the selected channel. In this example, the averaged adjacent energy parameter difference other than 0.1 is chosen as selected channels, which provides a discriminative feature map.

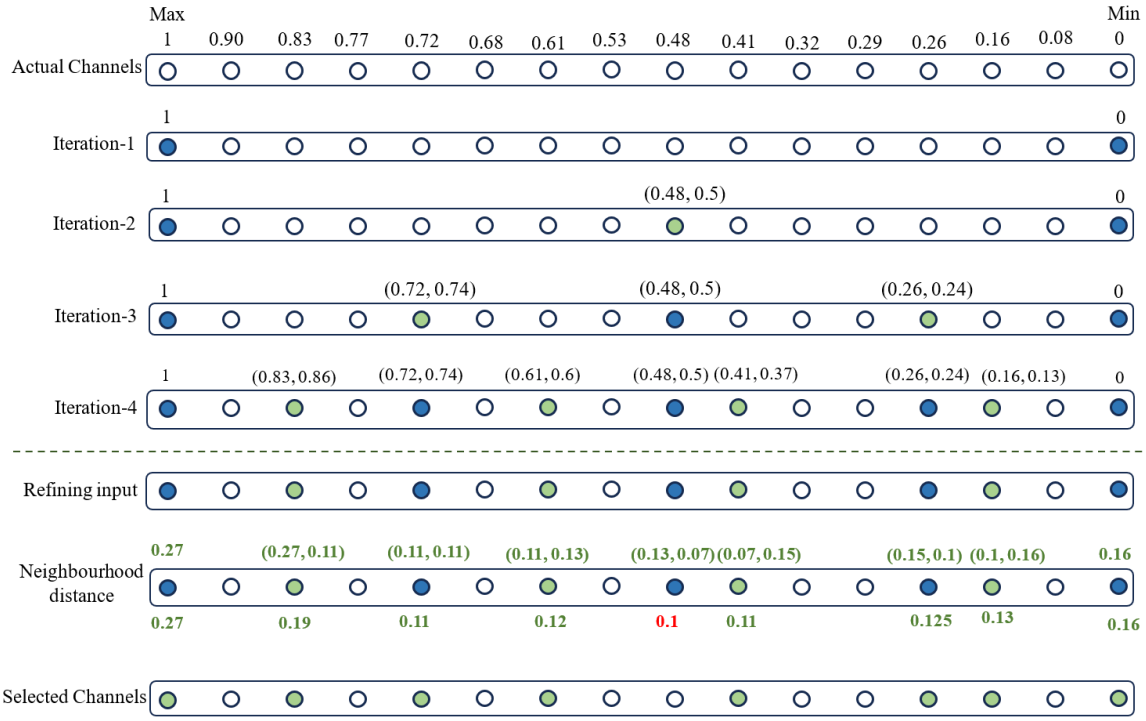


Fig. 2: Example representing the channel selection process with $R = 16$

The algorithm for the iterative channel selection process can be summarized as follows.

Algorithm 1:

Input: Descriptors with R channels (D_i), where $i = 1,2 \ldots \ldots R$

Output: Selected descriptor channels $\widehat{D}_i$ ($R/2$ channels).

Step 1: Compute the energy parameter δ_i for each feature map channel using the relation (4)

Step 2: Sort the energy parameter δ_i to obtain $\hat{\delta}_i$ and normalize the sorted energy parameters $\hat{\delta}_i$ to obtain σ_i.

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Step 3: Select the channels that correspond to the minimum (0) and maximum value (1) of σ_i in iteration 1.

Step 4: Select the channel whose sorted energy parameter is near to the average of energy parameter (i.e, 0.5) in iteration 2.

Step 5: Select the channel whose sorted energy parameter is near to the average of the energy parameter obtained between the minimum (0) (right side) and the energy parameter for the channel obtained in step 4. Similarly, select the channel whose sorted energy parameter is near to the average of energy parameters obtained between maximum (1) (left side) and the energy parameter for the channel obtained in step 4. This step will select two channels.

Step 6: Similarly, select the channels whose sorted energy parameter is close to the energy parameters corresponding to the channels selected in Step 5, with preceding steps (Steps 3 and 4).

Step 7: Repeat steps 3 to 6 till the condition $N_{s,j} > R/2$.

Step 8: Refine the selected channels by computing the average of the nearest difference between the energy parameters of the selected channels (left side, right side). i.e. the top $R/2$ channels that have a higher difference must be selected, while other channels must be eliminated. Thus, the selected channels contain the discriminative feature information $\widehat{D}_i$.

(ii) Gaussian Distribution-Based Spatial Attention

The Gaussian distribution-based spatial attention is performed on the channels that are selected by the channel selection module, which have $R/2$ channels that are selected from the descriptors. Let $\widehat{D}_i$ resembles the descriptors that are selected by the channel selection block. The GD-SA module will initially generate an attention map using the Gaussian distribution [25] and apply the attention map to the feature map. This will give more weightage to the regions that have more dominant descriptors and give less weightage to the regions that have less dominant descriptors. The attention map is computed on the region of interest; therefore, the scaled segmented lesion region mask $\widehat{M}$ is initially multiplied by the feature map $\widehat{D}_i$ to nullify the background region. For performing the GD-SA, two parameters, namely the intensity parameter and the gradient parameter, are computed on each point of the feature map using the 4-neighbourhood values. A 3D-Gaussian distribution is then constructed such that the normalized intensity parameter is used in one dimension, while the normalized gradient parameter is used in another dimension. The intensity parameter is computed as

$$\lambda_1(x,y) = \mu_1 - \frac{1}{5}\sum_{l=-1}^{1}\sum_{m=-1}^{1}\widehat{D}(x-l, y-m) \quad (6)$$

Here μ_1 resembles the average intensity computed on the feature map $\widehat{D}(x,y)$. Similarly, the gradient parameter can be computed as

$$\lambda_2(x,y) = \mu_2 - \tau_2(x,y) \quad (7)$$

Where $\tau_2(x,y)$ can be represented as

$$\tau_2(x,y) = \sum_{l\in\{-1,1\}}\sum_{m\in\{-1,1\}}\{\widehat{D}(x,y) - \widehat{D}(x-l, y-m)\} \quad (8)$$

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In equation (7) μ_2 can be obtained from the average of $\tau_2(x,y)$ computed on the feature channel of $\widehat{D}(x,y)$. The intensity and gradient parameters $\lambda_1(x,y)$ and $\lambda_2(x,y)$ are normalized between -1 and 1 based on the maximum magnitude.

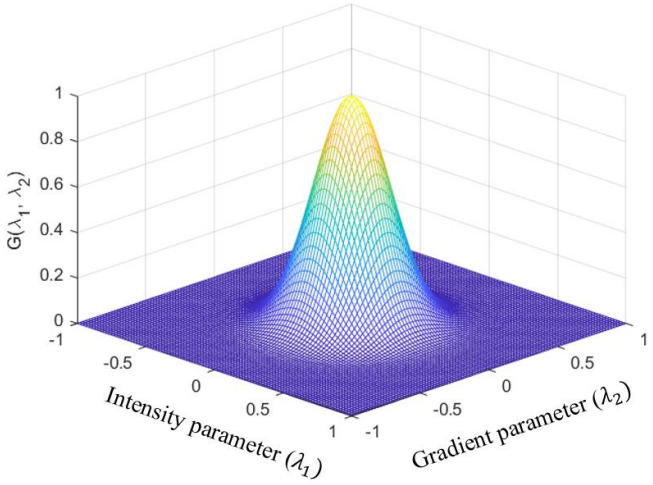


Fig. 3: Representation of 3D-Gaussian distribution in terms of intensity and gradient parameters

Let the normalized intensity parameter and gradient parameter be represented as λ_1 and λ_2 respectively. Thus, the Gaussian distribution can be expressed in terms of the parameters λ_1 and λ_2 be represented as,

$$G(\lambda_1,\lambda_2) = \exp\left(-\frac{{\lambda_1}^2+{\lambda_2}^2}{2\tau^2}\right) \quad (9)$$

Here τ resembles the standard deviation to control the Gaussian distribution. The output of the GD-SA can be mathematically expressed as,

$$\hat{S}(x,y) = \widehat{M}(x,y) \times G\big(\lambda_1(x,y),\lambda_2(x,y)\big) \times \widehat{D}(x,y) \quad (10)$$

$$\hat{S}(x,y) = \widehat{M}(x,y) \times \exp\left(-\frac{{\lambda_1}^2+{\lambda_2}^2}{2\tau^2}\right) \times \widehat{D}(x,y) \quad (11)$$

The function $G\big(\lambda_1(x,y),\lambda_2(x,y)\big)$ is a non-learnable parameter. Hence $\hat{S}(x,y)$ depends only on the input feature map $\widehat{D}(x,y)$.

(iii) Channel Duplication and Channel Elimination

The channel duplication and channel elimination are performed to maintain a uniform channel dimension for weighted fusion. The channel duplication duplicates the feature map, which increases the number of channels, while the channel elimination discards a few channels, which reduces the number of channels. For example, if the number of channels to the input of the channel duplication block is 8 and the number of channels after duplication is 32, then the channel duplication block will duplicate each channel four times. Fig. 4(a) illustrates the channel duplication process for converting 4 channels to 16 channels. The GDSA-U-Net architecture will have two channel duplication blocks where one converts 8 channels to 32 channels, while the other converts 16 feature channels to 32 channels.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

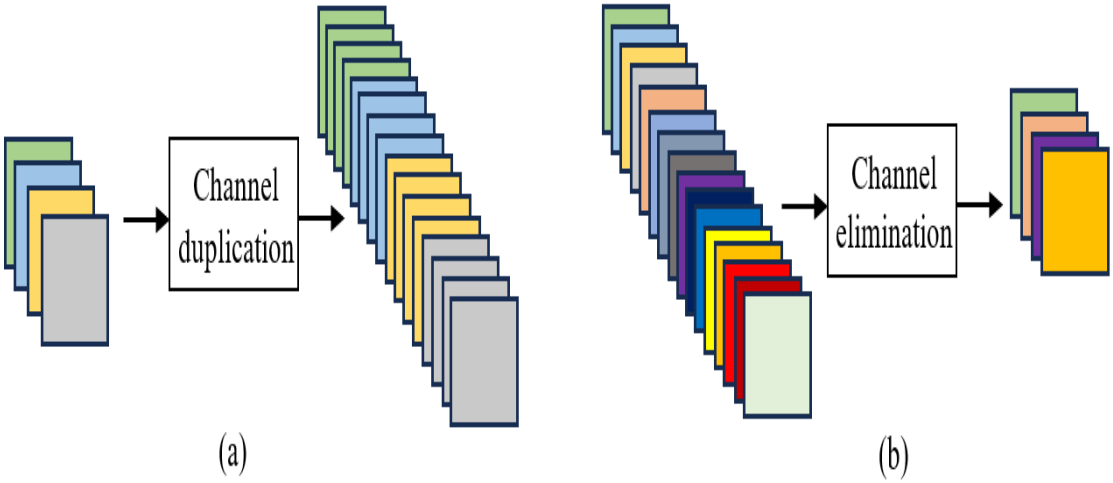


Fig. 4: Representation of Channel duplication and channel elimination (a) Channel duplication for converting 4 channels to 16 channels (b) Channel elimination

Similarly, in the case of channel elimination, where there is a need to convert 16 channels to 4 channels, the channel elimination block will eliminate three channels for every four channels. For example, if the number of channels input to the channel elimination block is 32 and the number of output channels required is 8, then the channel elimination block will eliminate 3 channels for every 4 channels. Fig. 4(b) illustrates the channel elimination process for converting 16 feature channels to 4 channels. The GDSA-U-Net architecture will have two channel elimination blocks, where one converts 64 channels to 32 channels, while the other converts 128 feature channels to 32 channels.

2.4 Weighted Fusion and Dense Network

Let H_1, H_2, H_3, H_4 and H_5 resembles the feature map input to the weighted fusion process that has 32 channels, having a spatial dimension of 64×64. The weighted fusion process includes five weights represented as w_1, w_2, w_3, w_4 and w_5. Thus, the output of the weighed fusion network can be expressed as,

$$H = \frac{w_1H_1+w_2H_2+w_3H_3+w_4H_4+w_5H_5}{w_1+w_2+w_3+w_4+w_5} = \frac{\sum_{k=1}^{5} w_kH_k}{\sum_{k=1}^{5} w_k} \qquad (12)$$

The weights w_1, w_2, w_3, w_4, and w_5 are scalars that get updated during the learning process. The output of weighted fusion also has a dimension of 32 channels with a spatial dimension of 64×64. The feature H is flattened using a flattening layer to obtain a one-dimensional feature vector having a size of 1×139264. The flattened feature is fed as the input to the dense network that has 7 or 8 output classes. The dense layer can able to differentiate the descriptors into one of the 7 or 8 classes, namely Basal cell carcinoma (BCC), vascular lesion (VAS), actinic keratoses (AKI), Squamous cell carcinoma (SCC), dermatofibroma (DEF), melanoma (MEL), melanocytic nevi (NV), and Benign keratosis (BKL).

3. EXPERIMENTAL RESULTS

The evaluation of the GDSA-U-Net approach was done utilizing the HAM-10000 [26] and ISIC-2019 [27] datasets. For evaluating the lesion classification process, evaluation scales, namely accuracy, MCC, Precision, F1-score, specificity, and Kappa, are used.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

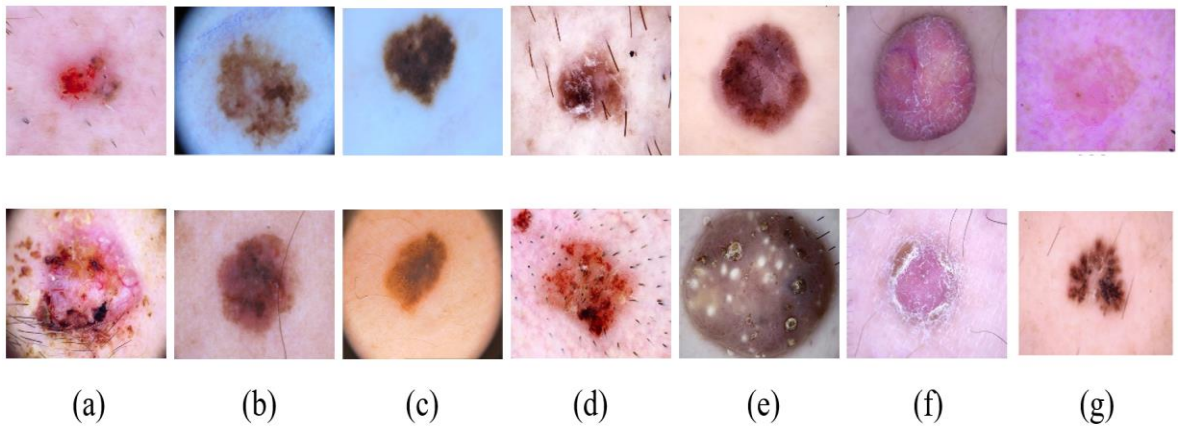


Fig. 5: Sample images used for the analysis of the GDSA-U-Net approach (a) BCC, (b) Mel, (c) NV, (d) VAS, (e) AKI, (f) BKL, (g) DEF, and (h) SCC

The HAM-10000 dataset has 10015 dermoscopic images with 7 classes, namely BCC, MEL, NV, VAS, AKI, BKL, and DEF. The dataset, ISIC-2019 dataset, has 16603 dermoscopic images with 8 classes, namely BCC, MEL, NV, VAS, AKI, BKL, DEF, and SCC. Fig. 5 shows the sample images used for analysis.

Table 1: Number of images used for analysis

Dataset	Augmentation	BCC	MEL	NV	VAS	AKI	BKL	DEF	SCC	Total
HAM-10000	Before	514	1113*	6705**	142	327	1099*	115	-	10015
	After	3598	1113*	3352**	994	2289	1099*	805	-	16603
ISIC-2019	Before	3323*	4522*	12875**	253	867	2624*	239	628	25331
	After	3323*	4522*	6438**	1771	6069	2624*	1673	4396	37253

The images from the HAM-10000 dataset are augmented with modifications such as brightening by 30, 60, darkening by 30, 60, and rotation by 90º, 270º. Table 1 depicts the number of images before and after augmentation. Here * resembles the classes in which no augmentation is performed, and ** resembles the classes in which only 50% of the images present in the dataset are used for analysis. The suggested GDSA-U-Net structure was trained with a batch size of 32, an initial learning rate of 0.0001, a maximum epoch of 100, with the Adam optimizer. The U-net uses an encoder depth of 4. The CLAHE algorithm uses a clip limit of 0.005 and a tile size of 8×8. The images are resized to 256×256 before training and testing. For the training of the GDSA-U-Net structure, 70% of the augmented images are utilized for training, and the remaining 30% are consumed for testing. The number of dermoscopic pictures in each category for training and testing the GDSA-U-Net is illustrated in Fig. 6.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

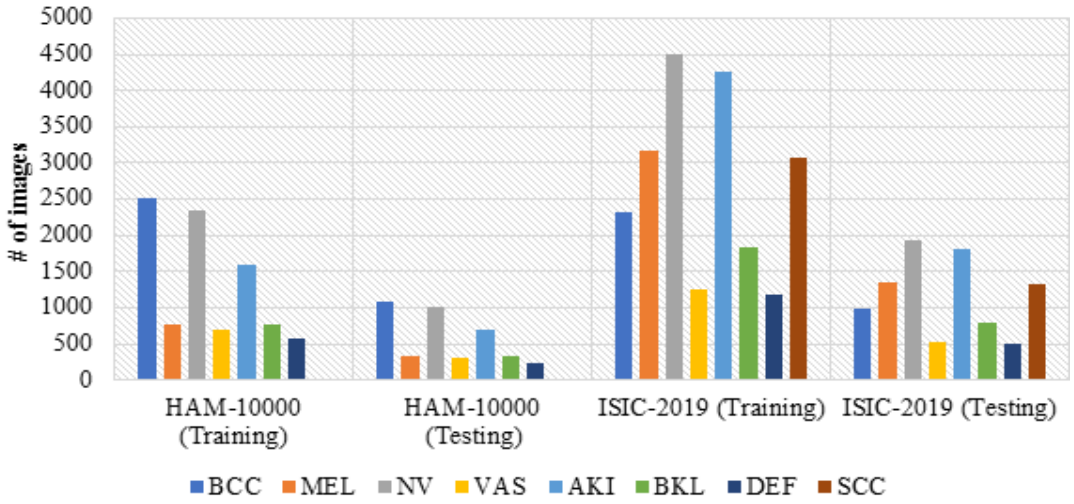


Fig. 6: Illustration of data distribution used for training and testing

The performance of the skin lesion segmentation results was evaluated using the datasets, namely Matthews correlation coefficient (MCC), Jaccard score index (JSI), dice score (DS), accuracy (Acc), Specificity (Spe), and Sensitivity (Sen).

Table 2: Performance comparison of skin lesion segmentation results obtained in the two datasets

Datasets	**MCC**	**JSI**	**DS**	**Acc**	**Spe**	**Sen**
HAM-10000	83.76	80.83	87.24	94.03	94.68	89.84
ISIC-2019	84.97	81.13	88.79	94.86	95.02	90.62

Table 2 provides the segmentation result obtained by the GDSA-U-Net. The segmentation network of the suggested GDSA-U-Net results in a JSI, DS, and accuracy of 80.83%, 87.24%, and 94.03%, respectively, when evaluated using the HAM-10000 dataset. Similarly, the segmentation network of the suggested GDSA-U-Net results in a JSI, DS, and accuracy of 84.97%, 81.13%, and 88.79%, respectively, when evaluated using the ISIC-2019 dataset. The segmentation results illustrate that the model can better segment the lesion region. A few of the sample segmentation result obtained by the segmentation network is presented in Fig. 7.

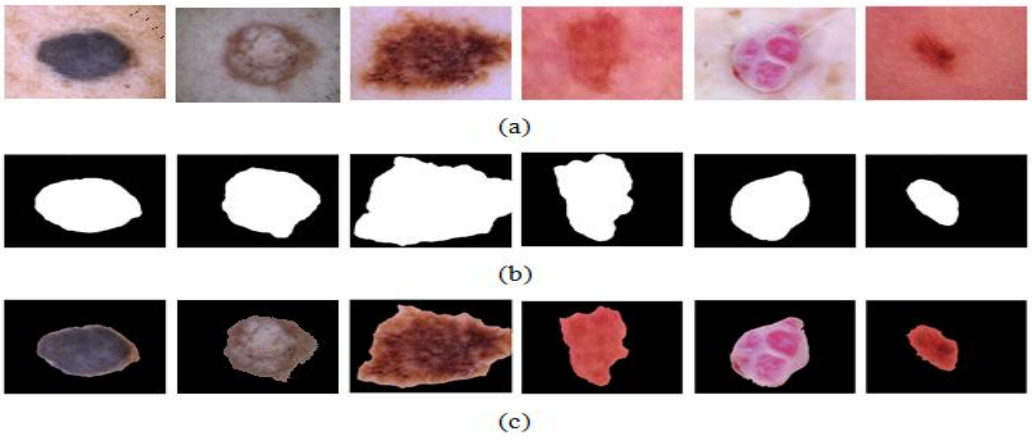


Fig. 7: Skin lesion region detected by the segmentation network (a) Input dermoscopic images, (b) Ground truth for lesion region, (c) Region segmented by the segmentation network

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 3: Comparison of classification evaluation measures obtained on different skin lesions when evaluated using the HAM-10000 dataset

Metrics	BCC	MEL	NV	VAS	AKI	BKL	DEF
Acc (%)	96.85	92.22	97.02	86.91	94.61	89.09	80.08
MCC (%)	95.49	89.15	95.55	87.66	92.65	88.86	82.62
Pre (%)	96.58	88.00	96.35	90.24	93.26	90.46	87.33
F1 (%)	96.71	90.06	96.68	88.55	93.93	89.77	83.55
Spe (%)	98.72	98.85	98.75	99.24	98.57	99.15	99.25

Table 3 provides the classification performance obtained by the suggested GDSA-U-Net for different skin lesions when analysed using the HAM-10000 dataset. For the lesion types BCC, MEL, NV, VAS, AKI, BKL, and DEF, the suggested GDSA-U-Net results in a classification accuracy of 96.85%, 92.22%, 97.02%, 86.91%, 94.61%, 89.09%, and 80.08%, respectively. The highest classification accuracy is obtained for the class BCC, and the lowest classification accuracy is obtained for the class DEF. For the class BCC, the suggested GDSA-U-Net approach yields a classification accuracy, MCC, precision, F1-score, and specificity of 96.85%, 95.49%, 96.58%, 96.71%, and 98.72%, respectively.

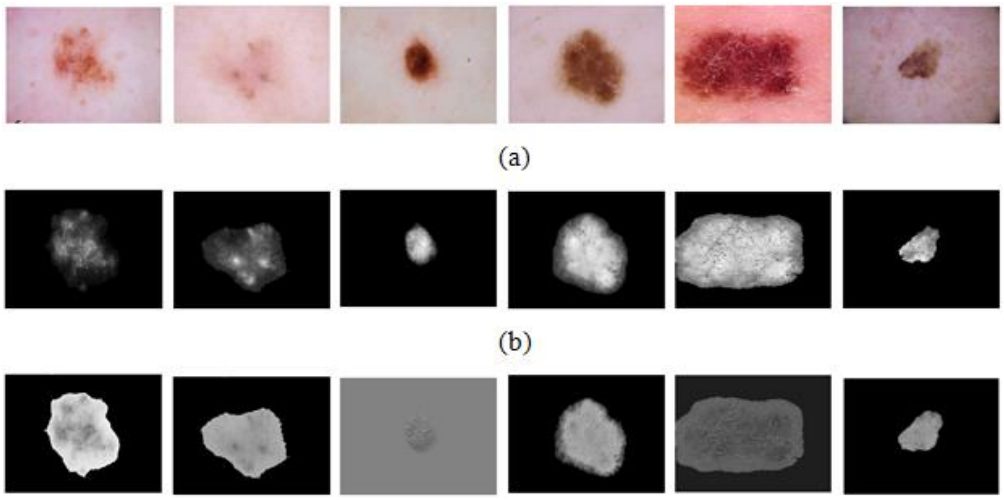


Fig. 8: Feature map obtained by the GD-SA attention (a) Input skin image (b) Gaussian distribution-based spatial attention map (c) Sample feature map

The feature map computed by the suggested Gaussian distribution-based spatial attention process is presented in Fig. 8. It shows the spatial attention map computed by the GD-SA. In this spatial attention map, the intensity near 0 provides low weightage, while the intensity near 1 provides a high weightage. Thus, the proposed GD-SA computes the dominant descriptor, neglecting the redundant feature region.

Table 4: Comparison of evaluation measures on different skin lesions when evaluated using the ISIC-2019 dataset

Metrics	BCC	MEL	NV	VAS	AKI	BKL	DEF	SCC

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Acc (%)	94.98	95.43	96.79	89.45	96.43	93.01	88.25	94.92
MCC (%)	94.16	94.36	95.28	89.54	95.65	92.95	88.52	94.16
Pre (%)	94.61	94.94	95.75	90.82	96.59	94.09	90.04	95.06
F1 (%)	94.79	95.19	96.27	90.13	96.51	93.55	89.13	94.99
Spe (%)	99.35	99.13	98.87	99.45	99.16	99.46	99.44	99.18

The analysis provided in Table 4 illustrates the performance obtained by the suggested GDSA-U-Net for different skin lesions when evaluated using the ISIC-2019 dataset. For the lesion types BCC, MEL, NV, VAS, AKI, BKL, DEF, and SCC, the suggested GDSA-U-Net results in a classification accuracy of 94.98%, 95.43%, 96.79%, 89.45%, 96.43%, 83.01%, and 94.92%, respectively. The highest classification accuracy is obtained for the class NV, and the lowest classification accuracy is obtained for the class DEF. For the class NV, the suggested GDSA-U-Net approach yields an accuracy, MCC, precision, F1-score, and specificity of 96.79%, 95.28%, 95.75%, 96.27%, and 98.87%, respectively.

Table 5: Performance comparison between the suggested GDSA-U-Net and traditional schemes when evaluated using the HAM-10000 dataset

Schemes	**Acc**	**MCC**	**Pre**	**F1**	**Spe**	**Kappa**
Ensemble DCNN [28]	90.23	85.65	87.22	86.31	95.05	89.26
Hierarchical fusion [29]	90.74	86.63	87.68	88.35	94.49	88.87
Swin transformer [30]	90.88	87.53	88.73	88.98	96.27	89.22
Tripple attention [31]	91.88	88.25	89.92	89.25	96.23	90.87
MS-fusion [32]	92.38	89.12	90.57	90.25	97.50	90.44
TI-ResNet [33]	92.76	90.00	90.71	90.08	97.77	91.89
MR-attention [19]	93.10	89.77	91.33	90.77	98.23	92.17
Proposed	93.71	90.28	91.75	91.32	98.93	92.73

The skin lesion classification performance of the suggested GDSA-U-Net approach was compared with the traditional schemes, namely MR-attention [19], TI-ResNet [33], MS-Fusion [32], Triple attention [31], Swin transformer [30], Hierarchical fusion [29], and Ensemble DCNN [28]. Table 5 provides a comparison between the proposed approach with these approaches when evaluated using the HAM-10000 dataset. The proposed approach results in an accuracy, MCC, precision, F1-score, specificity, and Kappa of 93.71%, 90.28%, 91.75%, 91.32%, 98.93%, and 92.73%, respectively.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

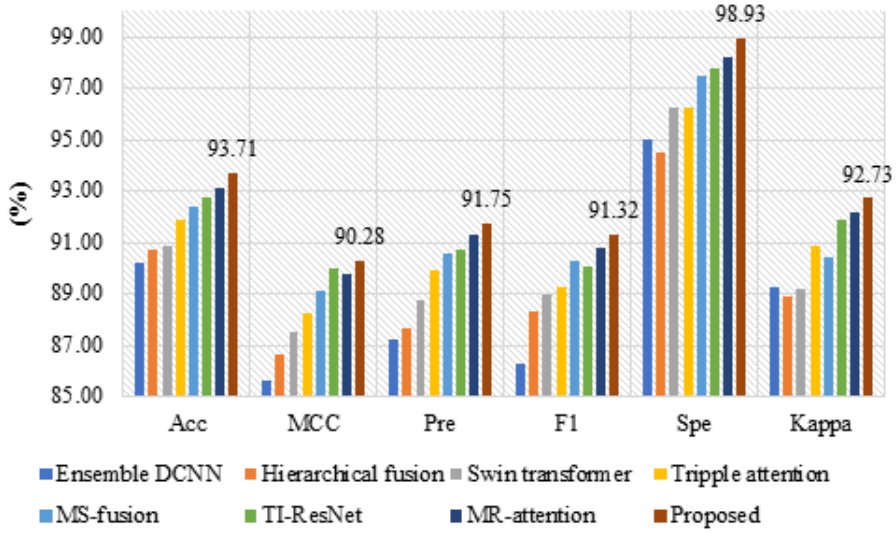


Fig. 9: Graphical comparison between the suggested GDSA-U-Net and traditional schemes when evaluated using the HAM-10000 dataset

The suggested approach results in an increase in accuracy, MCC, precision, F1-score, specificity, and Kappa of 0.61%, 0.51%, 0.42%, 0.55%, 0.70%, and 0.56%, respectively, then the MR-Attention approach. Further, the accuracy, MCC, precision, F1-score, specificity, and Kappa improve by 0.95%, 0.28%, 1.04%, 1.24%, 1.16%, and 0.84% than the TI-ResNet approach. The graphical comparison obtained when analysing using the HAM-10000 is illustrated in Fig. 9.

Table 6: Performance comparison between the suggested GDSA-U-Net and traditional schemes when evaluated using the ISIC-2019 dataset

Schemes	Acc	MCC	Pre	F1	Spe	Kappa
Ensemble DCNN [28]	90.48	89.42	90.21	90.34	95.72	90.59
Hierarchical fusion [29]	90.89	89.81	91.04	90.48	96.34	90.36
Swin transformer [30]	92.04	89.76	90.85	90.54	96.46	91.76
Tripple attention [31]	92.33	90.80	91.79	91.24	96.73	92.35
MS-fusion [32]	93.60	90.90	92.26	92.54	97.14	92.90
TI-ResNet [33]	93.14	92.13	92.74	93.24	98.11	93.54
MR-attention [19]	94.14	92.56	93.32	93.11	98.54	93.82
Proposed	94.85	93.08	93.99	93.82	99.25	94.14

Table 6 provides a comparison between the proposed approach with these approaches when evaluated using the ISIC-2019 dataset. The proposed approach results in an accuracy, MCC, precision, F1-score, specificity, and Kappa of 94.85%, 93.08%, 93.99%, 93.82%, 99.25%, and 94.14%, respectively.

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

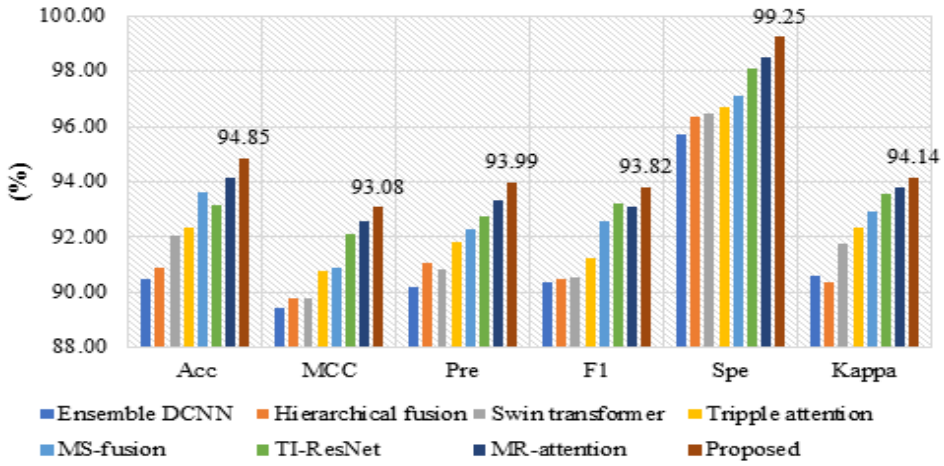
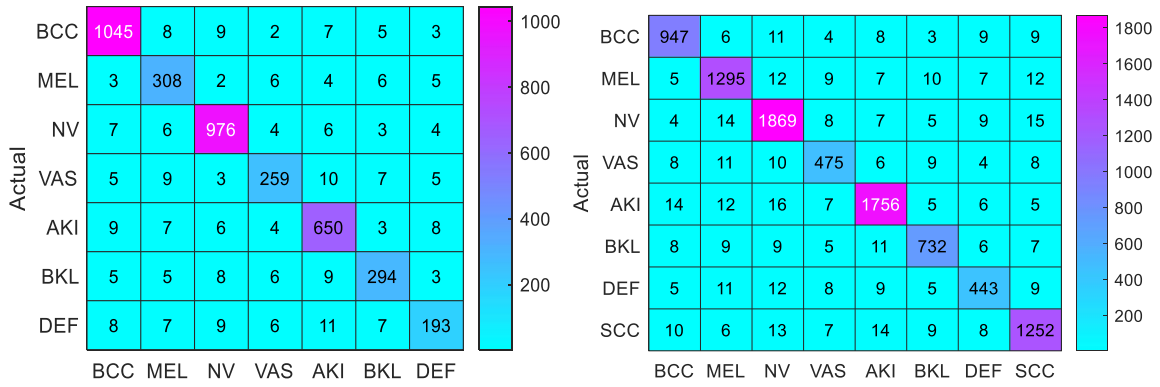


Fig. 10: Graphical comparison between the suggested GDSA-U-Net and traditional schemes when evaluated using the ISIC-2019 dataset

The suggested approach results in an increase in accuracy, MCC, precision, F1-score, specificity, and Kappa of 0.71%, 0.52%, 0.67%, 0.71%, 0.71%, and 0.32%, respectively, than the MR-Attention approach. Further, the accuracy, MCC, precision, F1-score, specificity, and Kappa improve by 1.71%, 0.95%, 1.25%, 0.58%, 1.14%, and 0.60% than the TI-ResNet approach. The graphical comparison obtained when analysing using the ISIC-2019 dataset is illustrated in Fig. 10.



(a) (b)

Fig.11: Confusion matrices obtained during the classification of Skin lesions by the GDSA-U-Net approach using the two dermoscopic image databases (a) HAM-10000 (b) ISIC-2019

The confusion matrices obtained during the classification of skin lesions by the GDSA-U-Net approach are illustrated in Fig. 11. Since the HAM-10000 dataset contains seven classes of skin lesions, namely BCC, MEL, NV, VAS, AKI, BKL, and DEF, the confusion matrix illustrates only seven classes, while the ISIC-2019 confusion plot also provides the class SCC lesion type. The confusion matrices obtained for the datasets HAM-10000 and ISIC-2019 datasets are presented in Fig.11(a) and Fig. 11(b), respectively. The confusion matrix illustrates that the suggested approach results in more true positives, which shows the classification performance. When comparing the classification performance obtained for the two datasets, the GDSA-U-Net approach when evaluated using the ISIC-2019 dataset results an increase in accuracy, MCC, precision, F1-score,

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

specificity and Kappa of 1.14%, 2.80%, 2.24%, 250%, 0.32% and 1.41% respectively than the HAM-10000 dataset.

4. CONCLUSION

This paper proposed a skin lesion detection and classification approach that uses two networks, namely the segmentation network and the classification network. The segmentation network uses the traditional U-Net structure, while the classification network uses the descriptors tapped from different sections of the U-Net encoder. Different other modules, namely channel selection, Gaussian distribution-based spatial attention, Channel elimination, and Channel duplication, are used in the classification network. The suggested GD-SA was derived using the Gaussian distribution that gives more weightage to the regions that have dominant descriptors, and this spatial attention is a non-learnable parameter. In addition to the GD-SA module, channel selection module, channel elimination, and channel duplication network are present in the classification network. The evaluation of the suggested GD-SA algorithm was done using the classification measures, namely accuracy, MCC, precision, and F1-score, using the HAM-10000 and ISIC-2019 datasets. The suggested GDSA-U-Net-based lesion classification approach results in an accuracy and Kappa value of 93.71% and 92.73% when categorizing 7 classes of lesions when analyzed utilizing the HAM-10000. Similarly, the suggested approach yields an accuracy and kappa value of 94.85% and 94.14%, respectively, when categorizing 8 classes of lesion types when analyzed using the ISIC-2019 dataset.

REFERENCES

[1]. Reichrath, J., Leiter, U., Eigentler, T., & Garbe, C. (2014). Epidemiology of skin cancer. *Sunlight, vitamin D and skin cancer*, 120-140.

[2]. Lomas, A. L. B. J., Leonardi-Bee, J., & Bath-Hextall, F. J. B. J. (2012). A systematic review of worldwide incidence of nonmelanoma skin cancer. *British journal of dermatology*, *166*(5), 1069-1080.

[3]. Kasparian, N. A., McLoone, J. K., & Meiser, B. (2009). Skin cancer-related prevention and screening behaviors: a review of the literature. *Journal of behavioral medicine*, *32*, 406-428.

[4]. Chinta, S. (2021). Advancements In Deep Learning Architectures: A Comparative Study Of Performance Metrics And Applications In Real-World Scenarios.

[5]. Krishnan, S. H., Vishwa, C., Suchetha, M., Raman, A., Raman, R., Sehastrajit, S., & Dhas, D. E. (2023). Comparative performance of deep learning architectures in classification of diabetic retinopathy. *International Journal of Ad Hoc and Ubiquitous Computing*, *44*(1), 23-35.

[6]. Zhang, J., Xie, Y., Xia, Y., & Shen, C. (2019). Attention residual learning for skin lesion classification. *IEEE transactions on medical imaging*, *38*(9), 2092-2103.

[7]. Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*, *86*, 25-32.

[8]. Qin, Z., Liu, Z., Zhu, P., & Xue, Y. (2020). A GAN-based image synthesis method for skin lesion classification. *Computer methods and programs in biomedicine*, *195*, 105568.

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[9]. Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A., & Wang, C. (2019). Fusing fine-tuned deep features for skin lesion classification. *Computerized Medical Imaging and Graphics*, *71*, 19-29.

[10]. Batista, L. G., Bugatti, P. H., & Saito, P. T. (2022). Classification of skin lesion through active learning strategies. *Computer Methods and Programs in Biomedicine*, *226*, 107122.

[11]. Hasan, M. K., Elahi, M. T. E., Alam, M. A., Jawad, M. T., & Martí, R. (2022). DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. *Informatics in Medicine Unlocked*, *28*, 100819.

[12]. Sathvika, V. B. T., Anmisha, N., Thanmayi, V., Suchetha, M., Dhas, D. E., Sehastrajit, S., & Aakur, S. N. (2024). Pipelined structure in the classification of skin lesions based on alexnet cnn and svm model with bi-sectional texture features. *IEEE Access*.

[13]. Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., ... & Xu, R. X. (2021). Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE transactions on medical imaging*, *41*(5), 1242-1254.

[14]. Sulthana, R., Chamola, V., Hussain, Z., Albalwy, F., & Hussain, A. (2024). A novel end-to-end deep convolutional neural network based skin lesion classification framework. *Expert Systems with Applications*, *246*, 123056.

[15]. Ding, S., Wu, Z., Zheng, Y., Liu, Z., Yang, X., Yang, X., ... & Xie, J. (2021). Deep attention branch networks for skin lesion classification. *Computer methods and programs in biomedicine*, *212*, 106447.

[16]. Zhang, D., Li, A., Wu, W., Yu, L., Kang, X., & Huo, X. (2024). CR-Conformer: A fusion network for clinical skin lesion classification. *Medical & Biological Engineering & Computing*, *62*(1), 85-94.

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

[18]. Akram, T., Alsuhaibani, A., Khan, M. A., Khan, S. U., Naqvi, S. R., & Bilal, M. (2024). Dermo-Optimizer: Skin Lesion Classification Using Information-Theoretic Deep Feature Fusion and Entropy-Controlled Binary Bat Optimization. *International Journal of Imaging Systems and Technology*, *34*(5), e23172.

[19]. Ghazouani, H. (2025). Multi-residual attention network for skin lesion classification. *Biomedical Signal Processing and Control*, *103*, 107449.

[20]. Shaheen, H., Singh, M. P., & Singh, A. K. (2025). Advanced skin lesion detection via efficientNetB0 and vision transformer model with spatial-aware attention. *Multimedia Tools and Applications*, 1-19.

[21]. Uthayakumar, G. S., Yaramadhi, M., Marimuthu, T., & Lakshmi, T. V. (2025). Optimized Mixed-Order Relation-Aware Recurrent Neural Networks based CAD Model for Skin Cancer Detection and Classification. *Knowledge-Based Systems*, *315*, 113222.

Received: August 08 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[22]. Setiawan, A. W., Mengko, T. R., Santoso, O. S., & Suksmono, A. B. (2013, June). Color retinal image enhancement using CLAHE. In International conference on ICT for smart society (pp. 1-3). IEEE.

[23]. Salido, J. A. A., & Ruiz Jr, C. (2017, June). Using morphological operators and inpainting for hair removal in dermoscopic images. In Proceedings of the computer graphics international conference (pp. 1-6).

[24]. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., ... & Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[25]. Zhang, Y., Fu, K., Wang, J., & Cheng, P. (2020). Learning from discrete Gaussian label distribution and spatial channel-aware residual attention for head pose estimation. *Neurocomputing*, *407*, 259-269.

[26]. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368.

[27]. Combalia, M., Codella, N. C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., ... & Malvehy, J. (2019). Bcn20000: Dermoscopic lesions in the wild. *arXiv preprint arXiv:1908.02288*.

[28]. Thurnhofer-Hemsi, K., López-Rubio, E., Dominguez, E., & Elizondo, D. A. (2021). Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting. *IEEE Access*, *9*, 112193-112205.

[29]. Hsu, B. W. Y., & Tseng, V. S. (2022). Hierarchy-aware contrastive learning with late fusion for skin lesion classification. *Computer methods and programs in biomedicine*, *216*, 106666.

[30]. Ayas, S. (2023). Multiclass skin lesion classification in dermoscopic images using swin transformer model. *Neural Computing and Applications*, *35*(9), 6713-6722.

[31]. Efat, A. H., Hasan, S. M., Uddin, M. P., & Mamun, M. A. (2024). A multi-level ensemble approach for skin lesion classification using customized transfer learning with triple attention. *PloS one*, *19*(10), e0309430.

[32]. Hu, Z., Mei, W., Chen, H., & Hou, W. (2024). Multi-scale feature fusion and class weight loss for skin lesion classification. *Computers in Biology and Medicine*, *176*, 108594.

[33]. Padhy, S., Dash, S., Kumar, N., Singh, S. P., Kumar, G., & Moral, P. (2025). Temporal Integration of ResNet Features with LSTM for Enhanced Skin Lesion Classification. *Results in Engineering*, 104201.

Received: August 08 2025