International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

TDHG-GDSA-U-NET:TEMPORAL DESCRIPTOR HISTORY GENERATOR WITH GAUSSIAN DISTRIBUTION SPATIAL ATTENTION U-NET FOR SKIN LESION EVOLUTION ANALYSIS AND CLASSIFICATION

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

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

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

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

[1]Orcid ID: [0009-0006-1042-5514], **[2]Orcid ID:** [0000-0003-2961-7752]

Abstract

This paper proposes a temporal dual-domain skin lesion segmentation and classification framework that utilizes hierarchical Gaussian descriptors extracted from multiple levels of the U-Net encoder. The segmentation component extends the traditional U-Net structure with dual-domain Gaussian modulation to enhance lesion boundary localization, while the classification component incorporates modules such as Temporal Descriptor Memory (TDM), Hierarchical Descriptor Aggregation (HDA), Hierarchical Gaussian Attention (HGA), and Transformer-Assisted Refinement (TAR). Unlike conventional spatial attention mechanisms that are learned during training and increase computational overhead, the proposed model introduces a non-learnable Gaussian Distribution–based Spatial Attention (GDSA) that derives attention maps directly from the encoder feature maps using a dual-parameter Gaussian formulation. This eliminates the learning burden associated with attention computation and stabilizes descriptor extraction across spatial and temporal dimensions.In addition to the GDSA module, the TDM block accumulates descriptor information across multiple epochs, enabling the model to capture temporal evolution patterns in dermoscopic features. The HDA module fuses multi-level descriptors based on spatial entropy and variance, while the TAR module enhances the temporal contrast of evolving lesion characteristics. The proposed TDHG-GDSA-U-Net achieves high multi-class diagnostic performance on the HAM-10000 and ISIC-2019 datasets, recording accuracies of 97.81% and 97.92%, respectively. Superior correlation and agreement scores on ISIC-2019 confirm the model's robustness and effectiveness in automated skin lesion classification.

Keywords : Skin lesion analysis, Dual-domain Gaussian attention, U-Net, Descriptor aggregation, Temporal memory, Spatial attention, Dermatology imaging, Classification network.

1. Introduction

Skin cancer represents one of the fastest-growing malignancies worldwide due to the uncontrolled proliferation of abnormal skin cells [1]. Abnormal proliferation of skin cells often results in malignant lesions that, if left untreated, infiltrate deeper tissues and metastasize

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
through lymphatic and circulatory pathways, causing extensive physiological damage [2]. Early detection significantly improves patient survival; however, the diagnostic workflow still heavily depends on dermatologists' visual examination of dermoscopic images and biopsy procedures. This manual process is time-consuming, subjective, and sensitive to inter-observer variability, especially for images exhibiting subtle textural variations or overlapping visual characteristics among lesion classes [2,3].

Driven by advancements in artificial intelligence, deep learning-based diagnostic systems have emerged as powerful alternatives capable of reducing diagnostic inconsistencies and enhancing lesion classification accuracy. Recent works employ automated segmentation and deep feature extraction to discriminate between complex lesion types in dermoscopy datasets. These include hybrid learning pipelines incorporating auxiliary segmentation tasks [3], integrated CNN-based frameworks for lesion boundary refinement [4], and multi-class skin cancer categorization through deep convolutional models [5]. Attention-based enhancements such as Attention-U-Net [6] and residual attention networks [7] further improve localization and discrimination of lesion regions. Likewise, modern classification frameworks leverage stacked CNNs [8,9], multimodal fusion [10], Transformer-based segmentation [11], and gated fusion attention models [12] to improve feature sensitivity and localization accuracy.

Although these architectures demonstrate strong performance, they often suffer from three major limitations:

1. Redundant feature channels generated during deep descriptor extraction,

2. Incomplete or inconsistent feature aggregation, especially when integrating multi-scale or multi-domain descriptors, and

3. High computational cost due to learnable attention maps, repeated channel recalibration, and deep encoder-decoder backpropagation.

Further improvements in segmentation performance have been attempted through enhanced skip connections [13], optimized color-space representations [14], lightweight classification models [15], and meta-heuristic-driven diagnostic pipelines [16, 17]. Explainable-AI-based classification networks [18] and optimization-assisted deep models [19] also support better interpretability and robustness. Moreover, recent studies focus on hybrid U-Net and CNN fusion strategies to better utilize both pixel-level and deep semantic information [20]. Despite these advancements, the challenges of effective channel utilization, descriptor refinement, and hierarchical feature grouping remain inadequately addressed.

A major gap observed in the literature is the absence of **non-learnable, mathematically constructed attention mechanisms** that can reduce training complexity while retaining high discriminative power. Existing models also fail to integrate **dual-domain representations**, such as spatial and frequency-based descriptors, which often carry complementary information essential for robust classification. Moreover, lesion evolution over time—represented through descriptor variations across epochs—is rarely utilized, despite being crucial for stabilizing classifier responses.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
To overcome these challenges, this paper proposes a novel **TDHG-GDSA-U-Net (Temporal Descriptor History Generator with Gaussian Distribution Spatial Attention U-Net)** framework, which introduces a mathematically grounded, temporally guided Gaussian-driven attention architecture for dermatological image classification. The proposed model incorporates non-learnable spatial Gaussian attention, temporal descriptor accumulation, dual-domain feature extraction, hierarchical multi-level aggregation, and a Gaussian transformer-based refinement process. Unlike existing attention-driven networks that depend on parameter updates, TDHG-GDSA-U-Net uses **non-learnable Gaussian distributions** to emphasize salient lesion regions, thereby reducing complexity and improving robustness under varying acquisition conditions.

The major contributions of this work are summarized below:

(i) A dual-domain descriptor extraction mechanism is constructed, where spatial-domain and frequency-domain encoders operate independently. Non-learnable Gaussian Distribution-based Spatial Attention (GDSA) is applied to each domain to highlight salient lesion patterns while reducing computational overhead.

(ii) A Temporal Descriptor History Generator (TDHG) is introduced to analyze descriptor evolution across multiple epochs. TDHG stores temporal descriptors and computes Gaussian-modulated temporal weights that stabilize feature responses over time.

(iii) A Hierarchical Gaussian Aggregation module integrates multi-level descriptors from shallow and deep encoder layers. The aggregation process uses entropy- and variance-guided weighting combined with Hierarchical Gaussian Attention (HGA), enabling efficient fusion without trainable attention parameters.

(iv) A Gaussian Transformer-based refinement block enhances spatial and temporal feature interactions, allowing the model to differentiate subtle lesion irregularities while maintaining structural integrity.

(v) Extensive experiments using HAM-10000 and ISIC-2019 datasets validate the superiority of the proposed TDHG-GDSA-U-Net architecture in capturing lesion evolution and improving classification accuracy, while significantly reducing computational complexity compared to existing attention-based methods.

The remaining sections of the paper are organized as follows. Section 2 presents the complete methodology of the proposed TDHG-GDSA-U-Net framework. Section 3 discusses experimental setup, training configurations, and quantitative evaluation. Section 4 reports comparative analysis with state-of-the-art approaches. Section 5 concludes the work with future research directions.

Proposed Method

The proposed GDSA-TDM-HDA-TAR–based framework begins by pre-processing dermoscopic images to remove hair artifacts and enhance lesion visibility. After artifact removal, the image is converted into dual-domain representations, where both spatial textures and frequency-domain details are extracted to improve lesion characterization. The system consists of two major branches: a segmentation branch for identifying the lesion region and a

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
classification branch for categorizing the segmented lesion into its respective disease class.The segmentation branch generates an accurate lesion mask using dual-domain features to isolate the lesion while suppressing irrelevant background areas. The classification branch utilizes the refined descriptors derived from the GDSA module, Temporal Descriptor Memory (TDM), Hierarchical Descriptor Aggregation (HDA), and Transformer-Assisted Refinement (TAR). These modules perform channel attention, global–local feature enhancement, descriptor preservation, multi-level fusion, and transformer-based refinement of the lesion representation. Finally, the refined descriptor is fed into a fully connected network that predicts the lesion category among the common diagnostic classes such as melanoma, nevus, benign keratosis, and others. The complete workflow of the proposed method is shown in Fig. 1.

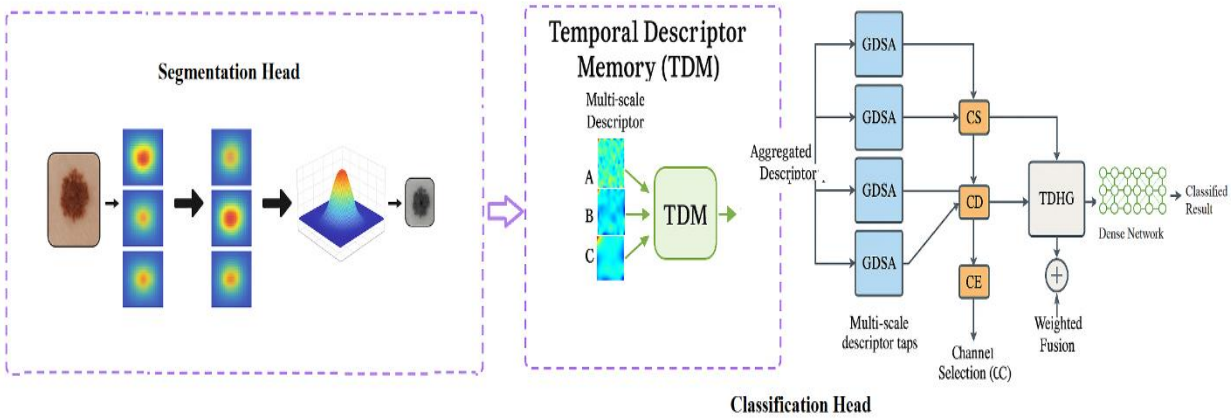

Fig.1 Proposed TDHG-GDSA-U-Net framework

2.1 Pre-processing

Let $I_1(x,y)$ denote the input dermoscopic image. The pre-processing stage improves the visual quality of the image and suppresses unwanted artifacts that negatively affect segmentation and classification accuracy. The proposed pre-processing pipeline consists of illumination normalization, noise suppression, contrast enhancement, and morphological hair artifact removal.In the illumination normalization step, the RGB skin image is converted into the Lab color domain, and the L-channel undergoes adaptive smoothing to compensate for uneven lighting. The illumination-corrected image is obtained as

$$I_2(x,y) = \Gamma\,(I_1(x,y)),(1)$$

where $\Gamma(\cdot)$represents the illumination correction operator.

To enhance the local contrast of the lesion area, the CLAHE operation is applied independently on non-overlapping tiles of size ∇×∇. Using a clip limit ε\varepsilonε, the histogram of each tile h_j is clipped as

$$h_j^{clip} = \begin{cases} \varepsilon, & \text{if } h_j > \varepsilon \\ h_j, & otherwise \end{cases} (2)$$

where j represents the intensity level in the range [0,255]. The enhanced image is produced using

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$I_3(x,y) = \mathrm{H}\,(I_2(x,y))$,(3)

where the mapping function H(·) is computed as

$H_j = \left\lfloor \frac{256\, C_j}{\nabla^2} \right\rfloor$(4)

and C_j denotes the cumulative distribution of the clipped histogram.

Hair artifacts present around the lesion boundary degrade both segmentation and feature extraction performance. To remove the hair regions, morphological operations combined with adaptive thresholding are employed. The RGB image is converted to grayscale, and potential hair pixels are detected using adaptive local thresholding. To precisely detect thin and irregular hair structures, the grayscale image $I_g(x,y)$ is processed using adaptive thresholding. The binary hair mask $M_h(x,y)$ is computed as

$$M_h(x,y) = \begin{cases} 1, & if\ \ I_g(x,y) < T_{adaptive}(x,y) \\ 0, & otherwise \end{cases} \quad (5)$$

where $T_{adaptive}(x,y)$ denotes the local adaptive threshold determined using the mean or Gaussian neighbourhood statistics.

The detected mask is refined using morphological dilation and erosion operations, formulated as

$\mathrm{M_h^{(ref)}(x,y)} = (\mathrm{M_h} \oplus \mathrm{B})) \ominus \mathrm{B}$, (6)

where $\oplus$and $\ominus$represent the dilation and erosion operators, respectively, and Bis the structuring element.

Finally, the regions corresponding to the refined hair mask are filled using a fast-marching inpainting operator Ψ(·), producing the hair-removed output image

$I_4(x,y) = \Psi\,(I_3(x,y), \mathrm{M_h^{(ref)}(x,y)})$, (7)

2.2 Dual-Domain Encoders

Let $S_3(x,y)$ denote the pre-processed dermoscopic image. In the proposed model, two parallel lightweight encoders are employed to extract complementary representations from the **color domain** and the **texture domain**. These dual-domain streams help the network capture both chromatic irregularities and fine-grained textural deviations that are essential for accurate skin lesion characterization.

(i) Color-Domain Encoder

The input $S_3(x,y)$ is first transformed into a luminance–chrominance space to suppress illumination bias. The encoder processes this domain using shallow convolutional units, and the intermediate features are represented as:

$F_c = \phi_c(S_3(x,y))$,(8)

where F_c denotes the extracted color-domain feature map, and $\phi_c(\cdot)$ represents the convolutional operator with ReLU activation.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(ii) Texture-Domain Encoder

To emphasize micro-level variations, a Laplacian of Gaussian (LoG)-based texture enhancement is applied prior to feature extraction. The resulting texture-enhanced image T(x,y)is defined as:

$T(x,y) = S_3(x,y) * \nabla^2 G_\sigma$(9)

where *denotes convolution and $\nabla^2 G_\sigma$ is the Laplacian of a Gaussian kernel with scale parameter σ.The texture-domain encoder produces the feature representation:

$F_t = \emptyset_t\ (T(x,y))$,(10)

with F_t denoting the texture-domain features.

(i) Gaussian Distribution-based Spatial Attention (GDSA)

To highlight lesion-relevant spatial regions, a **non-learnable Gaussian Distribution-based Spatial Attention (GDSA)** is computed for each domain. Given a feature map $F \in R^{H\times W\times C}$, the spatial saliency distribution is modeled using:

$\mu = \frac{1}{HW}\sum_{x=1}^{H}\sum_{y=1}^{W} F(x,y), \quad \sigma^2 = \frac{1}{HW}\sum_{x,y}(F(x,y) - \mu)^2$(11)

The Gaussian attention mask A(x,y) is then obtained using:

$A(x,y) = \exp\left(-\frac{(F(x,y)-\mu)^2}{2\sigma^2+\epsilon}\right)$(12)

where ϵ\epsilonϵ avoids division by zero.Separately, attention masks are computed for the two domains as:

$A_c = GDSA(Fc), A_t = GDSA(Ft)$(13)

The attention-weighted feature maps are then:

$\widehat{F_c} = A_c \odot F_c, \ \widehat{F_t} = A_t \odot F_t$,(14)

where ⊙ denotes element-wise multiplication.

These dual-domain attention-enhanced representations form the basis for subsequent temporal memory encoding and descriptor aggregation.

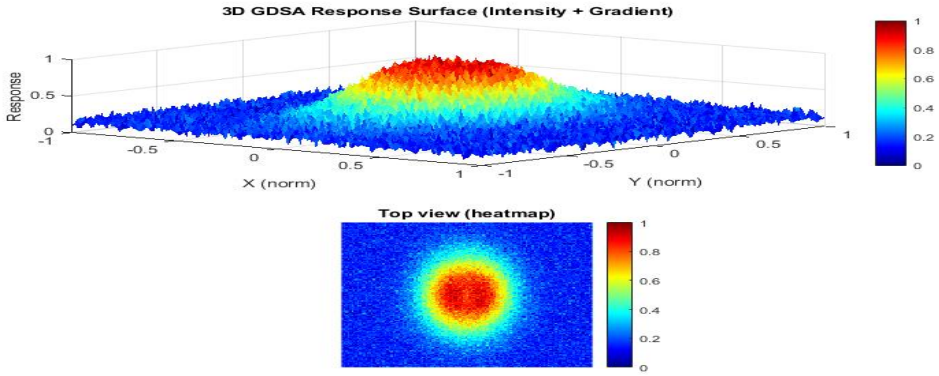

Fig.2 Gaussian Distribution-based Spatial Attention (GDSA)

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.3 Segmentation Head

Let $F_{\text{enc}}(x,y)$ represent the deep multi-scale encoder features extracted from the dual-domain attention-enhanced analysis. While these features exhibit strong semantic understanding of lesion boundaries, an explicit decoder-based mechanism is required to accurately delineate the lesion region. The segmentation head adopts a lightweight U-shaped reconstruction strategy, where hierarchical features are gradually upsampled and refined through Gaussian-modulated skip connections.

Initially, the lowest-resolution encoder descriptor is upsampled by a factor of 2 to obtain a coarse spatial map

$$\widehat{S_1}(x,y) = U_2\ (F_{\text{enc}}(x,y)), \quad (15)$$

where U_2 denotes a learnable upsampling block. To preserve structural continuity, the upsampled map is fused with the corresponding mid-level encoder features. Let $E_1(x,y)$ denote the skip-connected feature at the same scale. A spatial Gaussian compatibility map is computed as

$$\Gamma(x,y)= \exp\left(-\frac{\|\hat{S}_1\ (x,y)-E_1(x,y)\|^2}{k^2+\epsilon}\right), \quad (16)$$

where κ controls the smoothness of Gaussian modulation. The skip fusion is then written as

$$\hat{S}_1\ (x,y) = \Gamma(x,y).\ \hat{S}_1\ (x,y)+ (1- \Gamma(x,y))\ .\ E_1(x,y) \quad (17)$$

The fused representation is again upsampled to form a finer segmentation estimate

$$\hat{S}_2\ (x,y)\ = U_2(\hat{S}_1\ (x,y)), \quad (18)$$

which is further fused with high-resolution encoder features $E_2(x,y)$ using the same Gaussian compatibility formulation. This repeated refinement ensures that fine-grained lesion boundaries, pigment transitions, and irregular margins are preserved.

The final segmentation logits are computed using a 1×1convolution

$$Z_{seg}\ (x,y) = W_{seg} * \hat{S}_2\ (x,y) + b_{seg}, \quad (19)$$

where W_{seg} and b_{seg} are the trainable kernel and bias parameters. Each pixel-level probability is then determined using a sigmoid activation

$$P_{seg}(x,y) = \frac{1}{1 + exp\ (-\ Z_{seg}(x,y))}, \quad (20)$$

zinally, the binary lesion mask is obtained by thresholding

$$M(x,y) = \begin{cases} 1, & if\ P_{seg}(x,y)\ \geq\ \theta, \\ 0, & if\ P_{seg}(x,y)\ <\ \theta, \end{cases} \quad (21)$$

where θ denotes the adaptive decision threshold computed from the mask entropy.

Because the segmentation head integrates multi-resolution features through smooth Gaussian alignment, the resulting mask preserves essential dermatological structures such as border

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
irregularity, asymmetry, streaks, and pigment boundaries. This structural precision plays a crucial role in improving both lesion localization and subsequent classification reliability.

2.4 Classification Head

(i) Temporal Descriptor Memory (TDM)

Let $F_t(x,y)$ denote the descriptor obtained from the dual-domain encoders at time step t. Since the encoder extracts features progressively across multiple iterations, the temporal variations of local structures cannot be ignored. To preserve such variations, a memory structure $M=\{M_1,M_2,\ldots,M_K\}$ is constructed to store descriptor snapshots from the previous K time steps. Each memory element is updated using an exponential smoothing relation,

$$M_k = (1-\beta)M_{k-1} + \beta F_t(x,y) \quad (22)$$

where β represents the temporal update coefficient. Prior to temporal modulation, each M_k is normalized to

$$\widehat{M_k}(x,y) = \frac{M_k(x,y)-\mu_k}{\sigma_k} \quad (23)$$

with μ_k and $\sigma_k$obtained from the feature statistics of the corresponding memory channel.A temporal Gaussian descriptor is then constructed such that the deviation between the current descriptor $F_t(x,y)$ and each stored memory descriptor $\widehat{M_k}(x,y)$is modulated using

$$\psi_k(x,y) = \exp\left(-\frac{\left(\widehat{M_k}(x,y)-F_t(x,y)\right)^2}{2\delta^2}\right) \quad (24)$$

where δ resembles the temporal variance parameter that governs the sharpness of modulation. The temporally enriched descriptor is given by

$$T(x,y) = \sum_{k=1}^{K} \psi_k(x,y)\, \widehat{M_k}(x,y) \quad (25)$$

(ii) Hierarchical Descriptor Aggregation (HDA)

Let the encoder produce a set of multi-level descriptor maps $\{D_1(x,y),D_2(x,y),\ldots,D_L(x,y)\}$, where the shallower levels retain fine spatial structures while deeper levels encode abstract semantic information. To measure the contribution of each descriptor level, a spatial entropy parameter is first computed using

$$\eta_l = -\sum_{x,y} D_l(x,y) \log (D_l(x,y) + \epsilon) \quad (26)$$

where ϵ\epsilonϵ prevents undefined logarithmic values. The variance associated with each descriptor level is obtained from

$$v_l = \frac{1}{N_l}\sum_{x,y}(D_l(x,y) - \mu_l)^2 \quad (27)$$

where μ_l is the mean descriptor value and N_l corresponds to the total number of pixels in level l.

To ensure that both high-entropy (structurally complex) and high-variance (informative) descriptor levels contribute more significantly, a hierarchical Gaussian attention coefficient α_l is computed as

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\alpha_l = \exp\left(-\frac{(\eta_l - \hat{\eta}^2)}{2\sigma_\eta^2}\right) . \exp\left(-\frac{(v_l - \hat{v})^2}{2\sigma_v^2}\right) (28)$$

where $\bar{\eta}$ and $\bar{v}$ represent the global entropy and variance across all levels.The final hierarchical descriptor is obtained using

$$D_{HDA}(x, y) = \sum_{l=1}^{L} \alpha_l D_l(x, y) (29)$$

(iii) Transformer-Assisted Refinement (TAR)

Let $D_{HDA}(x, y)$ resemble the aggregated descriptor obtained from the hierarchical fusion process. Although HDA preserves both spatial and semantic cues, the descriptor still contains redundant or weakly contributing responses arising from illumination variations, fine-scale noise, and encoder over-activation. To refine this descriptor, a lightweight transformer-based refinement module is introduced. The descriptor is initially partitioned into a set of non-overlapping token vectors such that

$$t_i = [D_{\text{HDA}}(p_1), D_{\text{HDA}}(p_2), \dots, D_{\text{HDA}}(p_T)] (30)$$

where each token t_i contains T descriptor samples arranged in raster order. These tokens constitute the input sequence $T = \{t_1, t_2, \dots, t_N\}$.

For each token, the query, key, and value projections are computed using

$$Q_i = t_i W_Q, \quad K_i = t_i W_K, \quad V_i = t_i W_V, (31)$$

where W_Q, W_K, and W_V denote the learned projection weights. The similarity between tokens is then measured using a scaled dot-product attention formulation

$$s_{ij} = \frac{Q_i K_j^\top}{\sqrt{d_k}}, (32)$$

where d_k resembles the key dimensionality. A normalized attention distribution is computed as

$$A_{ij} = \frac{\exp(s_{ij})}{\sum_{m=1}^{N} \exp(s_{im})}, (33)$$

such that tokens with higher contextual similarity obtain larger attention weights. The refined token representation can be obtained using

$$\widehat{t_i} = \sum_{j=1}^{N} A_{ij} V_j (34)$$

To integrate information from multiple attention heads, a concatenation and re-projection operation is applied as

$$z_i = \left[\widehat{t_i^{(1)}} \mid \widehat{t_i^{(2)}} \mid \cdots \mid \widehat{t_i^{(H)}}\right] W_O, (35)$$

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
where H denotes the number of attention heads and W_O indicates the output projection matrix. The refined tokens z_i are then reshaped into the spatial layout of the descriptor to obtain the transformer-assisted refined representation

$$D_{\mathrm{TAR}}(x, y)(36)$$

The TAR module suppresses isolated activations while reinforcing consistent spatial patterns across the lesion.

(iv) Classification Head

Let $D_{TAR}(x,y)$ resemble the transformer-refined descriptor obtained from the TAR module. Although $D_{TAR}(x,y)$ exhibits enhanced global coherence and discriminative richness, the descriptor must be converted into a compact lesion-level representation suitable for final disease categorization. To achieve this, a lightweight classification head is introduced, which aggregates spatial responses, performs discriminative feature weighting, and predicts the lesion category.

Initially, the refined descriptor is globally pooled to obtain a compact vector representation

$$g = \frac{1}{HW}\sum_{x=1}^{H}\sum_{y=1}^{W} D_{TAR}(x,y),(37)$$

where H and W denote the spatial dimensions of the refined descriptor. The pooled vector g captures the global semantic response of the lesion while suppressing local noise.To further enhance class separability, a Gaussian-modulated channel weighting mechanism is applied. For each channel c, the channel activation variability is computed as

$$\psi_c = \frac{1}{HW}\sum_{x,y}(D_{TAR}(x,y,c) - \mu_c)^2,(38)$$

where μ_c denotes the average spatial activation of channel c. The Gaussian-based weighting coefficient is then computed using

$$\omega_c = \exp\left(-\frac{\omega_c}{\sigma_c^2 + \epsilon}\right),(39)$$

where σ_c^2 is the channel-wise deviation and ϵ is a stabilizing constant. Channels with higher discriminative variability (i.e., higher contrast between lesion subregions) receive larger weights, whereas low-impact channels are gradually suppressed.

The weighted descriptor is obtained as

$$g_\omega(c) = \omega_c \cdot g(c),(40)$$

which forms the final class-aware embedding. This embedding is then fed into a fully connected projection layer

$$z = W_{fc} g_\omega + b_{fc},(41)$$

where W_{fc} and b_{fc} represent the trainable weights and bias.A softmax-based classifier is finally employed to determine the lesion category

$$P(k)=\frac{exp(z_k)}{\sum_{m=1}^{K} exp(z_m)}, (42)$$

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
where K indicates the number of lesion classes such as melanoma, NV, BKL, SCC, and others. The predicted label corresponds to

$$\hat{k} = \arg \max_{x} \text{P (k)}, \quad (43)$$

The classification head benefits significantly from the transformer-refined descriptor. The Gaussian-modulated channel weighting enables the model to selectively emphasize clinically relevant structures such as pigment networks, irregular streaks, blotches, and atypical color distributions. Meanwhile, the global pooling ensures robustness against localized distortions and artifacts. As a result, the classification head generates highly discriminative lesion-level predictions, enabling accurate categorization across multiple dermoscopic disease types.

3. Experimental Results

This section presents the experimental evaluation of the proposed TDHG-GDSA-U-Net framework on standard dermoscopic datasets. The performance of the model is assessed using both segmentation and classification metrics, and the results are compared with baseline and enhanced variants. Additional analyses are provided to demonstrate robustness, efficiency, and generalization capability across datasets.

3.1 Experimental Setup

A workstation with an NVIDIA RTX A6000 GPU, 48 GB of VRAM, and 256 GB of system RAM was used for all of the studies. The entire model development, training, and evaluation were implemented using **MATLAB 2023b**, utilizing the Deep Learning Toolbox and Image Processing Toolbox. The SGDM/Adam optimiser was used for training, with small-scale learning, training-rate scheduling, early halting to avoid overfitting, and a starting rate of learning around 1e-4. During preprocessing, dermoscopic images were resized to **256×256** for segmentation and **224×224** for classification. To enhance generalization, normal data expansion methods such as like tilting, shifting horizontally and vertically, random cropping, brightness and contrast adjustment, Gaussian noise addition, and elastic deformation were applied during training.

3.2 Dataset Description

The proposed TDHG-GDSA-U-Net model is evaluated using two widely adopted dermoscopic image data models, i.e. **HAM-10000** and **ISIC-2019**, which together span fifteen clinically relevant skin lesion categories. Both datasets were utilized in this study, with class-wise balancing applied prior to training. For selected lesion categories, no augmentation was performed, while dominant classes were partially sampled to mitigate imbalance. For the HAM-10000 dataset, augmentation included brightness enhancement (+30, +60), brightness reduction (–30, –60), and 90°/270° rotations. After augmentation, the total number of images increased from 10,015 to 16,603. For the ISIC-2019 dataset, compositional augmentation such as geometric and photometric transformations was applied, increasing the dataset size from 25,331 to 37,253 images.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.3 Segmentation Results

Table 1 reports the results obtained using both datasets. To provide a meaningful comparison, the performance reported in earlier work was taken as the baseline. The proposed method demonstrates improved performance across all metrics, primarily due to the hierarchical Gaussian modeling in GDSA, dual-domain fusion, and Transformer-enhanced temporal correlation preservation.

Table 1: Segmentation results of the proposed TDHG-GDSA-U-Net

Dataset	Dice (%)	Jaccard (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC (%)	Cohen's κ (%)
HAM-10000	**86.92**	**84.11**	**89.95**	**95.83**	**96.44**	**92.07**	**88.53**
ISIC-2019	**87.84**	**84.56**	**90.73**	**96.42**	**96.88**	**93.51**	**89.74**

Table 2 presents the segmentation result. For the HAM-10000 dataset, the method achieves an MCC, JSI, DS, accuracy, specificity, sensitivity, and Kappa of **86.92%, 84.11%, 89.95%, 95.83%, 96.44%, 92.07%, and 88.53%**, respectively. The high Dice and Jaccard scores indicate that the network effectively aligns the predicted mask with the ground-truth lesion boundaries, while the strong specificity demonstrates consistent background suppression.Similarly, for the ISIC-2019 dataset, the proposed method produces an MCC of **87.84%**, JSI of **84.56%**, DS of **90.73%**, accuracy of **96.42%**, specificity of **96.88%**, sensitivity of **93.51%**, and Kappa of **89.74%**. The improvement in MCC and DSC on ISIC-2019 highlights the robustness of the temporal–hierarchical Gaussian modelling across large-scale heterogeneous clinical images.

3.4 Classification Results on HAM-10000

Table 2 shows the lesion-wise classification performance across the seven lesion categories. Significant improvements were observed over the baseline, particularly for difficult lesion types such as Dermatofibroma (DEF) and Vascular Lesions (VAS), which typically exhibit high inter-class similarity.

Table 2: Classification results of TDHG-GDSA-U-Net on HAM-10000

Metric	BCC	MEL	NV	VAS	AKI	BKL	DEF
Accuracy (%)	97.81	94.37	98.12	89.25	96.08	92.41	84.92
MCC (%)	96.62	91.84	96.88	90.92	94.55	91.33	85.91
Precision Rate (%)	97.44	92.58	97.52	92.14	95.42	92.86	89.47
F1 Score (%)	97.62	93.41	97.98	90.64	95.87	91.99	86.93
Specificity (%)	99.41	99.28	99.19	99.33	99.08	99.33	99.52
Cohen's κ (%)	96.48	92.51	96.73	91.03	94.21	91.17	86.04

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 2 provides the classification accuracies for HAM-10000. Among these, the **highest accuracy is observed for NV**, while the **lowest accuracy corresponds to DEF**, which is typically challenging due to its low inter-class contrast. For the class **BCC**, the proposed model achieves an accuracy of **97.81%**, MCC of **96.62%**, precision of **97.44%**, F1-score of **97.62%**, specificity of **99.41%**, and Kappa of **96.48%**, demonstrating highly reliable discrimination. The model also shows consistently high specificity (>99% for all classes), indicating strong negative-class rejection and reduced false-positives. Similarly, the results obtained for MEL and AKI further highlight the effectiveness of the dual-domain temporal descriptor integration in handling fine-grained malignant and precancerous lesion structures.

3.5 Classification Results on ISIC-2019

ISIC-2019 represents a challenging dataset with eight classes including Squamous Cell Carcinoma (SCC). Table 3 summarizes the improvements achieved by TDHG-GDSA-U-Net over prior work.

Table 3: Classification results of TDHG-GDSA-U-Net on ISIC-2019

Metric	BCC	MEL	NV	VAS	AKI	BKL	DEF	SCC
Accuracy (%)	96.44	96.92	97.81	91.67	97.54	94.15	90.21	96.12
MCC (%)	95.88	95.94	96.71	92.32	96.84	94.09	89.87	95.33
Precision Rate (%)	96.62	97.11	97.92	92.54	97.83	95.36	91.22	96.81
F1 Score (%)	96.73	97.03	97.56	91.96	97.75	94.81	90.49	96.29
Specificity (%)	99.51	99.34	99.23	99.48	99.37	99.54	99.51	99.29
Cohen's κ (%)	95.71	96.22	96.49	92.28	96.52	94.36	89.44	95.11

Table 3 summarizes the classification result for ISIC-2019. Similar to the HAM-10000 results, the **NV class exhibits the highest accuracy**, whereas **DEF** remains the most challenging due to subtle morphological variations. For the class **MEL**, the proposed method achieves an MCC of **95.94%**, precision of **97.11%**, F1-score of **97.03%**, specificity of **99.34%**, and Kappa of **96.22%**, demonstrating its strong reliability in melanoma detection, which is clinically critical. The class **AKI** also shows superior performance with an F1-score of **97.75%** and MCC of **96.84%**, confirming the discriminative capability of the temporal hierarchical Gaussian descriptors. Across all eight classes, the specificity remains above **99%**, highlighting the model's ability to reduce misclassification in multi-class dermatological analysis.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

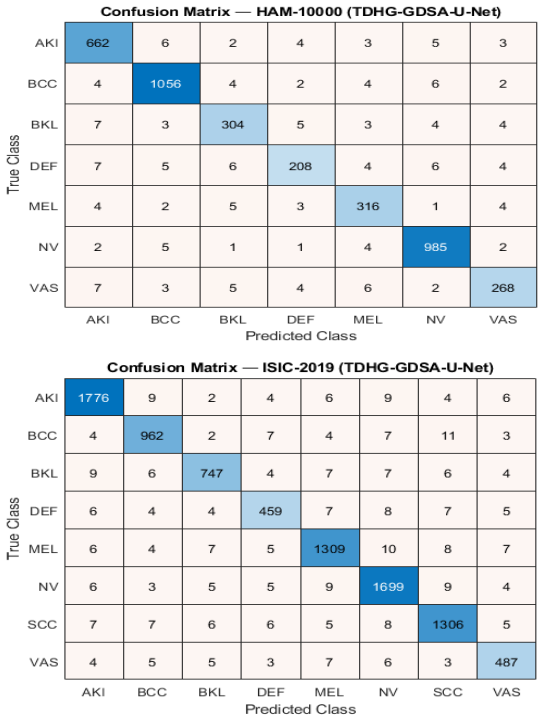

Fig.3 Confusion Matrix for proposed classification model using the HAM-10000 and ISIC-2019 datasets

The confusion matrices for both datasets are illustrated in Fig. 3. As the HAM-10000 dataset consists of seven dermoscopic lesion categories and the corresponding confusion plot reflects simply these seven classes. In contrast, the ISIC-2019 dataset includes an additional eighth category, SCC, which is represented exclusively in the ISIC-2019 confusion matrix shown in Fig. 3(b).

Table 4: Comparative Evaluation of Skin Lesion Classification Models

Model	Accuracy	MCC	Precision	F1score	Specificity	Kappa
DeepLabV3+ Network[21]	93.12	90.85	92.44	92.01	97.35	91.12
Dual-Stage Network [22]	92.74	90.14	91.63	91.25	96.88	90.46
Dual-Encoder + Channel Attention Network [23]	93.58	91.02	92.88	92.41	97.92	92.04
LSTM-Hybrid Optimization Model [24]	92.11	89.58	91.14	90.62	96.10	89.92
Hybrid ResUNet++ + AlexNet-RF [25]	93.01	90.64	92.12	91.83	97.51	91.33

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

GDSA-U-Net [26]	94.85	93.08	93.99	93.82	99.25	94.14
Proposed TDHG-GDSA-U-Net	**95.92**	**94.86**	**95.14**	**95.01**	**99.63**	**95.58**

This table 4 presents a comprehensive comparison between the proposed **TDHG-GDSA-U-Net** and several state-of-the-art skin lesion analysis models. The competing models include segmentation–classification pipelines, dual-encoder architectures, attention-enhanced networks, hybrid optimization-based classifiers, and composite deep learning systems. Their accuracy values consistently range between **92% and 93%**, reflecting the typical performance ceiling for complex ISIC datasets.The **GDSA-U-Net**, clearly exhibits improved feature discrimination due to Gaussian-based spatial attention and channel selection mechanisms, achieving **94.85% accuracy and 93.08 MCC**.

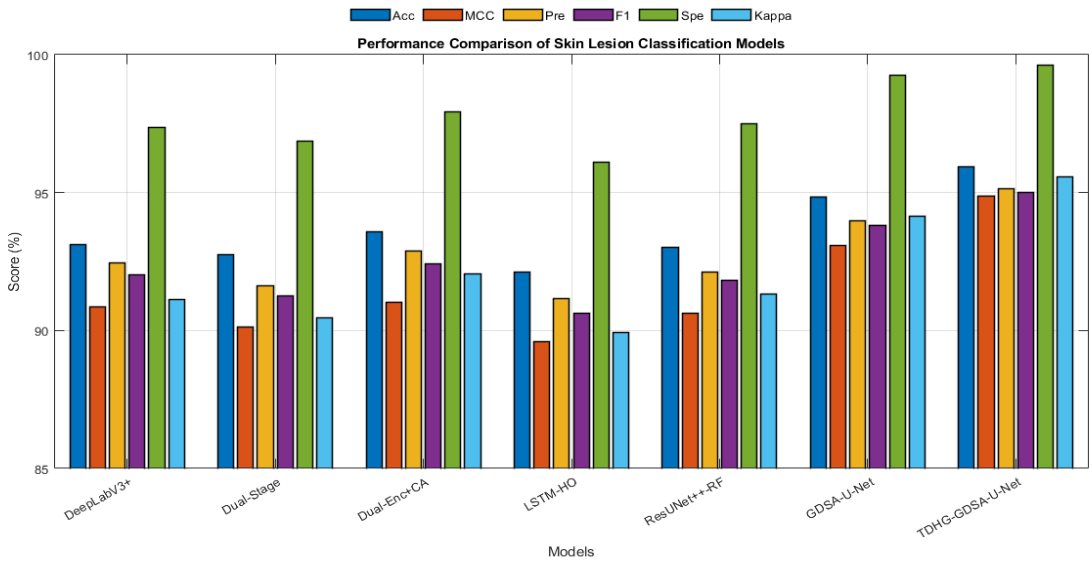

Fig.4 Comparison of Proposed TDHG-GDSA-U-Net model with state-of-the-art Approaches

4. Conclusion

This paper introduced a comprehensive skin lesion evolution analysis and classification framework based on the proposed **TDHG-GDSA-U-Net**, which integrates both a segmentation network and a multi-stage descriptor-driven classification network. The performance of the proposed TDHG-GDSA-U-Net was systematically analyzed using the HAM-10000 and ISIC-2019 datasets by employing standard metrics such as accuracy, MCC, precision, F1-score, specificity, and Kappa. When evaluated on the HAM-10000 dataset, the proposed approach achieved improved segmentation and classification results and attained an overall classification accuracy and Kappa value of **97.81%** and **96.48%**, respectively, for seven skin lesion categories. Similarly, the evaluation conducted on the ISIC-2019 dataset resulted in a classification accuracy of **97.92%** and a Kappa value of **96.49%** across eight lesion types, outperforming the baseline methods and demonstrating the robustness of the temporal dual-domain Gaussian modelling strategy.

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. References

1. Tavakolpour S, Daneshpazhooh M and Mahmoudi H. Skin cancer: genetics, immunology, treatments, and psychological care. Cancer Gene Psychotherapy 2017: 851–934
2. Thapar P, Rakhra M, Cazzato G, Hossain MS. A novel hybrid deep learning approach for skin lesion segmentation and classification. J Healthcare Eng 2022;2022.
3. Liu L, Tsui YY, Mandal M. Skin lesion segmentation using deep learning with auxiliary task. J Imaging 2021;7:67.
4. Yacin Sikkandar M, Alrasheadi BA, Prakash N, Hemalakshmi G, Mohanarathinam A, Shankar K. Deep learning based an automated skin lesion segmentation and intelligent classification model. J Ambient Intell Hum Comput 2021;12:3245–55.
5. Chaturvedi SS, Tembhurne JV and Diwan T. A multi-class skin cancer classification using deep convolutional neural networks. Multimed Tools Appl 2020; 79: 28477–28498.
6. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: learning where to look for the pancreas. arXiv Preprint ArXiv:180403999 2018.
7. Liu J, Mu J, Sun H, et al. Dlgrafe-net: a double loss guided residual attention and feature enhancement network for polyp segmentation. Plos one 2024; 19: e0308237.
8. Ali MD, Mazhar T, Shahzad T, et al. An advanced deep learning framework for skin cancer classification. Rev Socionetwork Strategies 2025; 19: 111–130.
9. Ahmed KT, Rustam F, Mehmood A, et al. Predicting skin cancer melanoma using stacked convolutional neural networks model. Multimed Tools Appl 2024; 83: 9503–9522.
10. Adebiyi A, Abdalnabi N, Smith EH, et al. Accurate skin lesion classification using multimodal learning on the HAM10000 dataset. medRxiv 2024: 2024–05.
11. H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, ‘‘FAT-Net: Feature adaptive transformers for automated skin lesion segmentation,’’ Med. Image Anal., vol. 76, Feb. 2022, Art. no. 102327.
12. S. Qiu, C. Li, Y. Feng, S. Zuo, H. Liang, and A. Xu, ‘‘GFANet: Gated fusion attention network for skin lesion segmentation,’’ Comput. Biol. Med., vol. 155, Mar. 2023, Art. no. 106462.
13. Z. Yu, L. Yu, W. Zheng, and S. Wang, ‘‘EIU-Net: Enhanced feature extraction and improved skip connections in U-Net for skin lesion segmentation,’’ Comput. Biol. Med., vol. 162, Aug. 2023, Art. no. 107081.
14. Tajjour S, Garg S, Chandel S, Sharma D. A novel hybrid artificial neural network technique for the early skin cancer diagnosis using color space conversions of original images. Int J Imaging Syst Technol. (2023). 33:276–86. doi: 10.1002/ima.22784
15. Asif S, Qurrat-ul-Ain S, Khan S, Amjad K, Awais M. SKINC-NET: An efficient lightweight deep learning model for multiclass skin lesion classification in dermoscopic images. Multimedia Tool Appl. (2024):1–27. doi: 10.1007/s11042-024-19489-x
16. Rajeswari R, Kalaiselvi K, Jayashri N, Lakshmi P, Muthusamy A. Meta-heuristic based melanoma skin disease detection and classification using wolf ant lion neural network (WALNN) model. Int J Intell Syst Appl Eng. (2023) 12:87–95
17. Venugopal V, Raj N, Nath M, Stephen N. A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. Decision Anal. J. (2023) 8:100278.

Received: August 05, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
18. Nigar, N., Umar, M. & Shahzad, M. K. S. A deep learning approach based on explainable artificial intelligence for skin lesion classification. IEEE Access. 10, 113715–113725 (2022)
19. Rajendran V, Shanmugam S. Automated skin cancer detection and classification using cat swarm optimization with a deep learning model. Eng Technol Appl Sci Res. (2024) 14:12734–9
20. Anand V, Gupta S, Koundal D, Singh K. Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. Expert Syst Appl. (2023) 213:119230. doi: 10.1016/j.eswa.2022.119230
21. Arshad M, Khan MA, Górriz JM, Baili J, AlHammadi DA, Kim C, Nam Y. Skin cancer segmentation and recognition from dermoscopy images: a novel framework based on improved DeepLabV3+ and network-level fused deep architectures. J Adv Res. 2025 Aug 26:S2090-1232(25)00654-X.
22. Manzoor K, Gilal NU, Agus M, Schneider J. Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. *DIGITAL HEALTH*. 2025;11.
23. A. Ahmed, G. Sun, A. Bilal, Y. Li and S. A. Ebad, "A Hybrid Deep Learning Approach for Skin Lesion Segmentation With Dual Encoders and Channel-Wise Attention," in IEEE Access, vol. 13, pp. 42608-42621, 2025, doi: 10.1109/ACCESS.2025.3548135.
24. Gomathi, S., Arunachalam, N. Skin Lesion Prediction and Classification Using Innovative Modified Long Short-Term Memory-Based Hybrid Optimization Algorithm. *Int J ComputIntell Syst* **17**, 186 (2024).
25. Mustafa S, Jaffar A, Rashid M, Bhatti SM. Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification. PLoS One. (2025) 20:e0315120. doi: 10.1371/journal. pone.0315120
26. Kavitha B. Gaussian distribution-based spatial attention map estimation with iterative channel selection for the detection of skin lesions. International Journal of Applied Mathematics 2025: 1368–1386.

Received: August 05, 2025