Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

An Accurate Multiclass Skin Lesions Classification of Benign and Malignant Using Deep Learning and Dermoscopic Analysis

Punam R Patil[1], Ritu Tandon[2]

[1]Research Scholar, [2]Associate Professor, Computer Science & Engineering, SAGE University, Indore, India.

E-mail: [1]punamrpatil78@gmail.com, [2]ritu.tondon@sageuniversity.in

Abstract

Skin cancer, the leading type of cancer, poses a serious risk to public health, most notably melanoma, which is fatal if not treated. Early diagnosis is essential, but traditional diagnosis has low precision due to the poor image quality and the challenges of visual discrimination. Using the publicly available HAM10000 dataset, which is suitable for segmentation and multi-class classification, we propose a robust deep learning-based hybrid system for classification, segmentation, and severity analysis of skin cancer. The model begins with advanced preprocessing in the ELPWF module to pre-process the image and eliminate noise. Enhanced images are then processed with the TS-HCaps feature extraction algorithm to capture complex temporal and hidden features while reducing the dimensionality problem. The best features are selected using the TCWOA module to reduce computational complexity before segmentation by the PA-HRST model, achieving an HD analysis of 4%, and an ASSD of 0.008078, which is higher than existing schemes. The extracted features are then forwarded to the GA-MSKAD hybrid classification model, employing global attention to extract channel and spatial features and accurately classify skin cancer types including AKIEC, BCC, BKL, DF, MEL, NV, and VASC with a 99.18% accuracy, precision of 99.09%, recall of 99.13%, specificity of 99.03% and an F1-score of 99.11%. Finally, the severity is forecast by applying the RLLM regression model with residual and lasso analysis to achieve RMSE of 0.282, MAE of 0.08, and MSE of 0.08. This complete approach from image enhancement and feature extraction to hybrid classification and severity analysis is far superior to conventional diagnostic techniques. To enhance interpretability, stability, and clinical practical utility, future studies will prioritize the inclusion of Explainable Al (EAI), multiple datasets, and clinical data.

Keywords: Skin Cancer, Capsule Network, Walrus Optimization, Swin Transformer, Segmentation, Deep Learning (DL).

1. Introduction

The skin is the body's first organ for prevention against infections, heat, and ultraviolet radiation. Skin cancer is among the most frequent cancers in the world. Early detection is vital as it profoundly enhances patient prognosis and survival rates. Nevertheless, conventional diagnostic methods can lead to misinterpretation due to their tendency to be often subjective,

DOI: https://doi.org/10.36548/jiip.2025.3.019
Received: 25.07.2025, received in revised form: 26.08.2025, accepted: 12.09.2025, published: 23.09.2025

time-consuming, and reliant on expert interpretation. Consequently, techniques of deep learning and artificial intelligence in dermatological diagnosis have been extensively developed and implemented. To address the issue of skin lesion classification, a lightweight dynamic kernel-based convolutional neural network (CNN) was presented. This model decreased the computational complexity typically related to deep learning algorithms, yet showed effective multiclass skin cancer classification [1]. In order to enhance melanoma recognition and classification accuracy, the authors integrated CNNs with the Aquila Optimizer to improve feature extraction and feature reduction [2].

Presenting a new and improved Mixed-Order Relation-Aware Recurrent Neural Network (MoR-RNN) for enhanced skin cancer diagnosis is another instance of sophisticated network design [3]. Their CAD approach generated higher classification accuracy through capturing elaborate spatial and sequential relationships in the dermoscopic images. Through their consideration of existing and novel therapeutic strategies for non-melanoma skin cancer, this review supports such computational endeavors and highlights the value of accurate early diagnosis in selecting the optimal treatment strategy [4].

Strong multiclass classification with a deep CNN model demonstrates its effectiveness on large dermoscopic datasets. Lesion classification under varied conditions was enhanced, largely due to their approach [5]. Additionally, a deep learning multiclass model known as DSCC_Net was suggested. It uses dermoscopic images to deliver high-accuracy diagnoses, greatly aiding in automated clinical decision-making [6]. EOSA-Net, an enhanced CNN framework designed for multiclass skin cancer detection, considerably improves classification reliability. To improve feature learning, their method uses convolutional enhancements [7]. By providing more high-resolution visual information, this novel AI-based super-resolution image reconstruction method aims to improve the quality of input images and support the early detection of skin cancers [8]. Also, by enhancing the feature extraction and classification processes, an AI-based classification model known as SkinNet was developed that improved the accuracy of skin cancer diagnosis [9].

It also demonstrates how deep CNNs can be utilized in practice to classify skin disease, proving the growing reliability of AI for dermatological use [10]. A robust deep learning design was developed to enhance the accuracy of multiclass classification in different situations, such as noise and unbalanced samples [11]. For automatic classification, specialised melanoma deep CNNs were employed to demonstrate that CNN-based solutions are feasible in real-world clinical environments [12]. With the application of deep learning to optical coherence tomography (OCT) images, the scope of AI diagnostics can be expanded beyond dermoscopy.

Their model showed the capability of deep learning in high-resolution cellular imaging by distinguishing between different types of skin cells [13]. A novel hybrid deep learning model has enhanced classification accuracy for various types of skin cancer through the use of several convolutional layers and attention mechanisms [14]. Lastly, a hybrid diagnostic approach incorporating attributes of multiple CNN models was outlined for the early detection of skin lesions. The significance of ensemble and feature fusion methods to improve classification generalization and robustness was presented in their work [15].

1.1 Motivation & Objectives of the Study

The numerous existing deep learning methods for skin cancer diagnosis have notable limitations that the proposed study aims to overcome. Many prior works suffer from high

computational complexity, making them less suitable for real-time or resource-constrained clinical environments. Some models improve feature extraction but still struggle with minute artifacts such as hair, moles, and bubbles, which negatively impact classification accuracy. Others achieve good performance but fail to generalize well in imbalanced datasets, noisy images, or variable clinical conditions. Moreover, conventional approaches often demand longer processing times, limiting their efficiency in practical healthcare use. To address these gaps, the proposed study introduces an advanced framework that combines improved feature extraction, optimal feature selection through novel optimization techniques, progressive attention mechanisms, and global semantic knowledge distillation. Together, these innovations reduce computation time, enhance robustness against artifacts, and significantly improve the accuracy and reliability of multiclass skin cancer detection. The primary contribution of the proposed model is examined in the following manner.

1. To introduce an improved feature extraction module for the classification of skin cancer based on a two-phase self-attention based Hierarchical Capsule Network (TS-HCaps).
2. To provide the Tent Chaotic Walrus optimization algorithm (TCWOA) within a detection framework that yields the best characteristics for classifying skin cancer while using less time.
3. To provide a multi-scale hierarchical residual swin transformer (PA-HRST) based on Progressive Attention for effective skin cancer segmentation.
4. To effectively classify skin cancer, we present a Global Attention-based Multilevel Semantic Knowledge Alignment Distillation Network (GA-MSKAD).
5. To predict the severity of skin cancer, a Residual Lasso Logistic Regression model (RLLR) is presented.

1.2 Organization of the Paper

This manuscript is divided into the following sections: With its problem formulation, motivation, and study objectives, Section 1 offers a thorough overview of deep learning methodologies relevant to skin cancer diagnosis and the objectives of the study, while Section 2 offers a thorough review of pertinent techniques related to skin disease identification. A brief summary of the methodology used for the skin cancer detection model suggested in this work is included in Section 3, along with illustrative images. A thorough evaluation of the performance measures attained by the proposed model in relation to current approaches is presented in Section 4, with the aid of graphical representations. The overall findings drawn from the suggested model are explained in Section 5, along with potential future directions.

2. Related Work

To improve diagnosis accuracy, recent developments in skin cancer detection have used a variety of deep learning and hybrid approaches. A survey of various related techniques for skin cancer segmentation and classification is presented below.

To enhance lesion border localization and classification [16] suggested a model that combined an attention-guided Capsule Network with Active Contour Snake. An improved [17] lesion analysis using a dynamic graph cut technique and a full-resolution CNN, In this [18] presented a distributed capsule neural network for scalable and effective identification of epidermal lesions. To extract both geographical and semantic data, [19] combined capsule networks with Graph Neural Networks. Also, an optimization using a genetic algorithm was addressed in [20].

Additional capsule-based advancements were introduced in [21] and [22], the latter of which combined a fuzzy logic-based F-CapsNet with adaptive fuzzy-GLCM segmentation. In [23] a dual encoder framework was proposed for accurate and effective segmentation. In this work [24] utilized a DSNET to offer deployable and lightweight models. Moreover, [25] used a modified attention mechanism to create a reliable CNN model. When taken as a whole, these studies show how contemporary skin cancer diagnostic systems prioritize spatial awareness, interpretability, and computing efficiency. The reviews of various related techniques for skin cancer segmentation and classification are given in Table 1.

Table 1. Overview of the Existing Models

Author Name & Reference, Year	Technique name	Performance	Demerits
Behara et al. [16], 2024	Active contour segmentation, ResNet50, Capsule Network, SGD	Accuracy-98% AUC-ROC-97.3%	Lack of transparency and interpretation in capsule activations.
Adla et al. [17], 2023	Dynamic graph cut algorithm, Transferring models, and FrCN architecture	Accuracy-97.98% Precision-94.33%	Depending solely on the data quality. Lack of interpretation.
Dubey et al. [18], 2024	GHO-Capsule Neural Network, SSFO, Hybrid GHO optimizer	Accuracy -99.06%, specificity-97.83% and sensitivity-99.50%	Difficulties in processing large amounts of clinical data.
Santoso et al. [19], 2024	GNN, Capsule network, Tiny Pyramid ViG	This model attained 95.52% accuracy after 75 epochs of training.	Depending solely on the specific dataset. Imbalanced datasets.
Salih et al. [20], 2023	CNN, DICE	Accuracy-98.66%	Reliance on quantity and quality of training data.
Sivasangeetha et al. [21], 2023	Coot search optimization, Capsule network	Accuracy-99.26	Limited and imbalanced datasets.
Ali et al. [22], 2023	GLCM, Capsule network, dynamic routine technique, F-CapsNet	99.16% (ISBI 2017), 99.45% (ISBI 2019), 98.42% (PH2)	Limited generalization analysis. Potential bias in the training data.

Ahmed et al. [23], 2025	DuaSkinSeg model, MoblieNet v2, and ViT-CNN	This model enhanced diagnosis performance.	Limited adaptability
Chen et al. [24], 2025	Lightweight DSNET model	MIoU – 80.23%, DSC – 89.06% (ISIC 2017); MIoU – 81.30%, DSC – 89.81% (ISIC 2018)	Lack of analysis of generalization performance.
Thwin et al. [25], 2025	DCAN-Net	Accuracy – 97.57%, Recall – 97.57%, F1 – 97.10%	Lack of transparency.

2.1 Research Gap

High accuracy is achieved by models such as those suggested by [16, 17] but they require a large amount of computing power, which may not be feasible in settings with limited resources. Developing more lightweight and efficient models that do not compromise accuracy remains a critical area for future research. Research focused on enhancing model performance with low-quality or non-standardized images could make these tools more universally applicable. Despite achieving high performance, models like the one suggested by [18] often function as "black boxes," making it challenging for clinicians to trust and accept these technologies. Enhancing the interpretability of these models could bridge the gap between AI researchers and medical practitioners.

3. Proposed Work

The well-ordered DL approaches proposed in this study are ideal for skin cancer segmentation and classification. Below Figure 1 illustrates the architectural design of the proposed framework. To improve feature accuracy, the data were first pre-processed for de-noising. An enhanced low pass wiener filter (ELPWF), which successfully eliminates noise while maintaining important image properties, was used for this pre-processing stage. Following noise reduction, the data are examined, and features are extracted using the TS-HCaps model, which effectively captures intricate temporal correlations and latent characteristics while focusing on the most crucial aspects of the image data through the use of self-attention techniques.

Complex temporal connections and latent features were effectively captured by TCWOA, which selects the retrieved features. Furthermore, TCWOA minimizes unnecessary computational overhead by ensuring that feature selection is performed efficiently. Following pre-processing, data are input into the PA-HRST model, which uses its hierarchical structure to more easily analyse the data while capturing both local and global aspects. This increases the transparency and explicability of a model's decisions. The GA-MSKAD approach uses distillation techniques to improve classification accuracy and global attention mechanisms to align multilevel semantic knowledge for SC classification. After categorization, the RLLR model was used to assess the severity of the identified cancer with different severity metrics.

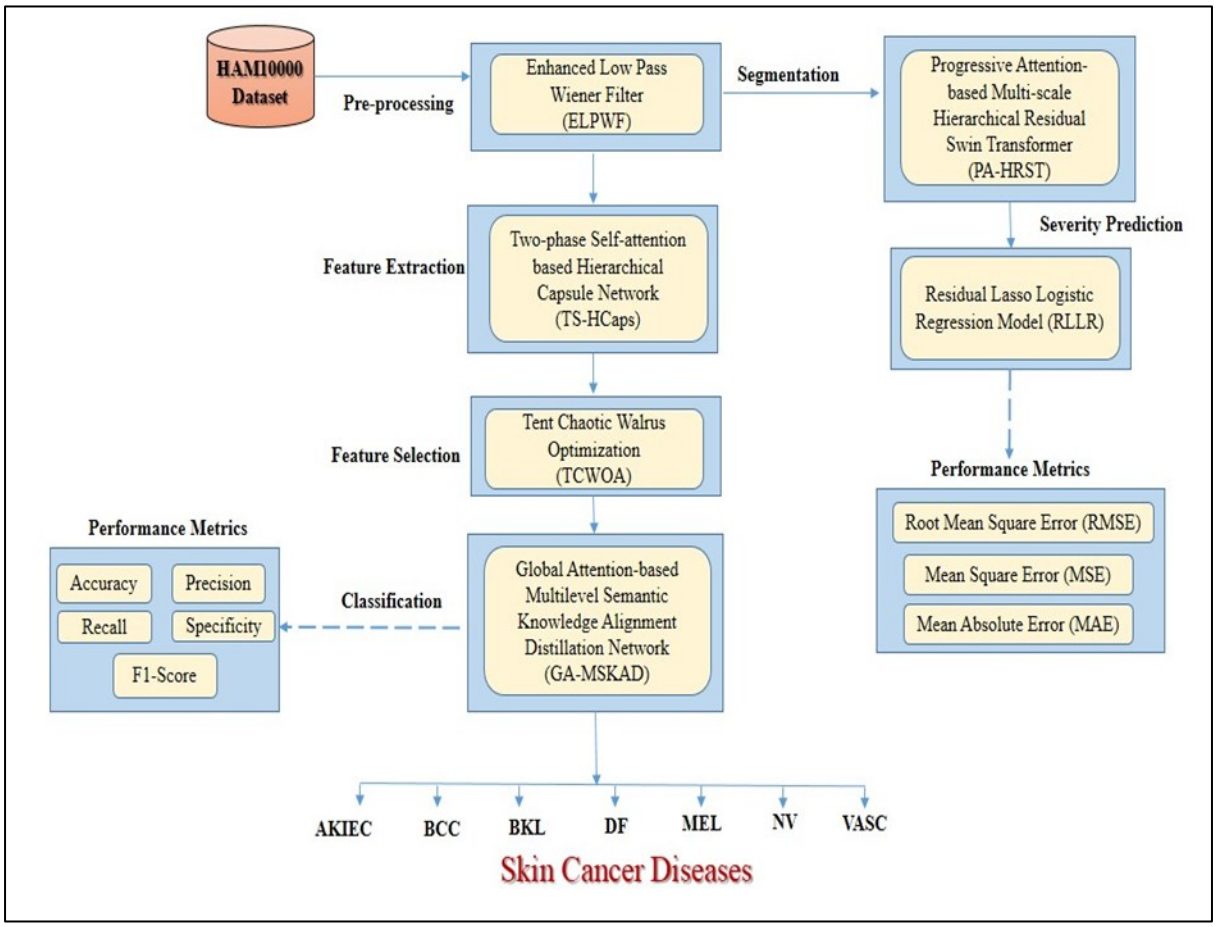


Figure 1. Architectural View of the Proposed Work

3.1 Enhanced Low-Pass Wiener Filter

A filter model is frequently used to improve image values by eliminating high-frequency noise, interpolating, and de-blurring. To guarantee that the retrieved features are useful, noise-free, and indicative of the actual properties of the lesion, image pre-processing is essential. In particular, the Wiener filter, an Enhanced Low-Pass Wiener Filter (ELPWF) is a popular pre-processing method for reducing image noise while maintaining small details such as lesion boundaries, color variation, and texture, all of which are critical for precise classification. The hybrid model, also referred to as the ELPWF model, was used in this study to process the following data. To smooth the actual continuous conditioning variables and reduce the significance of the aberrant values, a low-pass filter was first employed.

The low-pass filter is expressed as:

$$Y(n) = \sum_{k=0}^{k=N-1} h(k) \cdot x(n-k) \tag{1}$$

Where $h(k)$ denotes the filter coefficient, $x(n)$ is the input, and $Y(n)$ denotes the filtered output. The 2D Gaussian low-pass filter is given by:

$$G(x,y) = \frac{1}{2\pi\sigma 2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

The cut-off frequency is where the response drops to a maximum of 0.607. The Wiener filter is:

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)||H(u,v)|^2 + \frac{S_N(u,v)}{S_F(u,v)}}\right] G(u,v) \quad (3)$$

It reconstructed degraded images, reduced noise, and enhanced the accuracy of the proposed ELPWF model.

3.2 Two-Phase Self-Attention based Hierarchical Capsule Network

This approach uses the proposed TS-HCaps model with pre-processed data as input. An advanced deep learning architecture called the two-phase self-attention based hierarchical capsule network (TS-HCaps) was created to improve feature representation and classification accuracy in challenging image analysis tasks such as skin cancer lesion classification, where both global structure and local texture are essential. TS-HCaps integrates: self-attention mechanisms for adaptive feature weighting across spatial and channel dimensions, hierarchical feature learning for multi-scale abstraction, and capsule networks for modelling part-whole interactions, and is divided into two stages to distinguish between high-level and low-level decision learning.

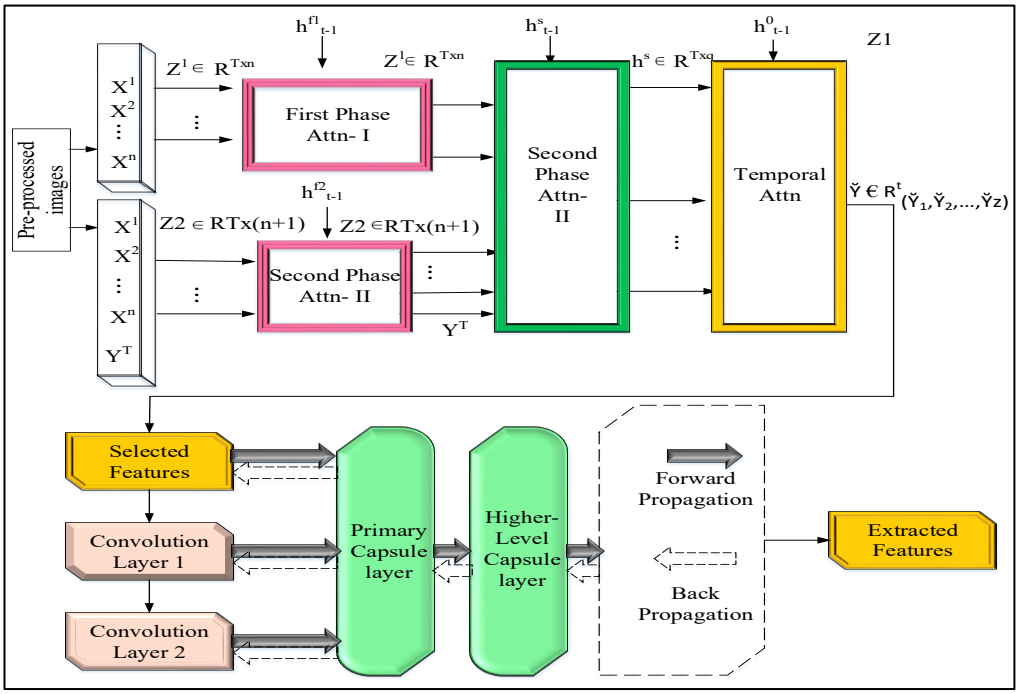


Figure 2. Architecture of TS-HCaps Model

The above Figure 2 illustrates the architecture of the TS-HCaps model. To improve feature extraction and categorization, the proposed architecture combines a capsule network with a multiphase attention mechanism. Prior to the Second Phase Attention (Attn-II), which improves contextual awareness by using parallel attention paths, pre-processed input data are subjected to First Phase Attention (Attn-I) to capture local interdependence. A Temporal Attention module, that records sequential dependencies over time, was used to further enhance

these outputs. The Primary Capsule Layer retains low-level characteristics and spatial linkages after the attention-enhanced data are processed through convolutional layers to extract spatial patterns. The Higher-Level Capsule Layer, which uses dynamic routing to describe part-whole linkages, further refines these layers. While backpropagation modifies the network to maximize performance, forward propagation through these capsule layers produces deeply extracted features. This hybrid technique is well suited for tasks requiring accurate categorization and temporal awareness because it combines the advantages of capsule networks for robust, hierarchical feature encoding with attention mechanisms for spatial-temporal focus.

3.3 Tent Chaotic Walrus Optimization Algorithm

Issues with earlier feature selection models, such as PSO [26], Grey Wolf Optimization (GWO) [27], and the Whale Optimization Algorithm (WOA) [28], include inadequate convergence rates, lack of optimal depth, and becoming trapped in local optima. The high detection rate in the WOA model led to a fast rate of convergence. This model creates a single flattened feature that is then fed into the TCWOA model by combining features, including color, size, texture, and skin tone contrast. The best buildings are selected by combining the WOA with a tent chaotic map. Table 2 shows the pseudocode of TCWOA for feature selection as follows:

Table 2. Pseudocode of TCWOA for Feature Selection

Algorithm: Pseudocode of TCWOA
Input: Algorithm parameter (population size M , maximum repetition S)
Produce Initialize population by tent map and define relevant parameter
Calculate fitness and achieve ideal solution
While $s \leq S$
  If $|Danger\_signal| \geq 1$ {Exploration stage}
    Modify fresh location of every walrus
  Else {Exploitation Stage}
    If $Safety\_signal \geq 0.5$
      For every male walrus
        Modify fresh location founded on Halton series
      End for
      For every female walrus
        Modify fresh position
      End for
      For every juvenile walrus
        Modify fresh location
      End for
    else
      If $|Danger\_signal| \geq 0.5$
        Modify fresh location of every walrus
      else
      End if

```
            End if
        End if
    Modify walrus location
    Evaluate fitness value and modify current ideal solution
      s = s + 1
End while
Outcomes ideal solution
```

An optimization method inspired by nature, the Tent Chaotic Walrus Optimization Algorithm (TCWOA) is useful for detecting skin diseases, especially during the feature selection stage. Many features are taken from photos or clinical data in medical image analysis, such as when classifying skin conditions like psoriasis, eczema, or melanoma. A lot of these elements might be unnecessary or redundant, which could hurt diagnostic models effectiveness and performance. This is addressed by TCWOA, which increases model accuracy and lowers computing costs by choosing the most informative subset of characteristics. The algorithm imitates walrus behavior, balancing exploration and exploitation through adaptive mechanisms. A tent map is used to initialize the population, increasing diversity and preventing premature convergence. Candidate solutions ("walruses") update their positions based on simulated danger and safety signals, enabling efficient search-space navigation. By integrating chaotic dynamics and role-based behavior, TCWOA avoids local optima, enhances robustness, and ensures search diversity. Beyond medical applications, it also proves useful in engineering, finance, and other optimization-driven fields.

3.4 Global Attention-based Multilevel Semantic Knowledge Alignment Distillation Network

In skin disease detection, large dermoscopic images contain numerous features related to color, texture, shape, and borders, making optimal feature selection essential to identify the most informative attributes. Attention mechanisms, specifically channel and spatial attention, further enhance these features by focusing on critical spectral channels and spatial regions, such as lesion borders or irregularities, resulting in refined features that improve classification accuracy. These enhanced features are then input into two models: a deep, complex teacher model that achieves high accuracy but is computationally intensive, and a lighter, faster student model designed for practical use in mobile or real-time applications. Through knowledge distillation, the student model learns by mimicking the teacher model's outputs and internal representations, enabling it to attain comparable performance with reduced complexity and faster inference times. Global attention produces improved features- that are fed into the MSKAD model to classify skin cancer. Figure 3 illustrates the architecture of the GA-MSKAD model.

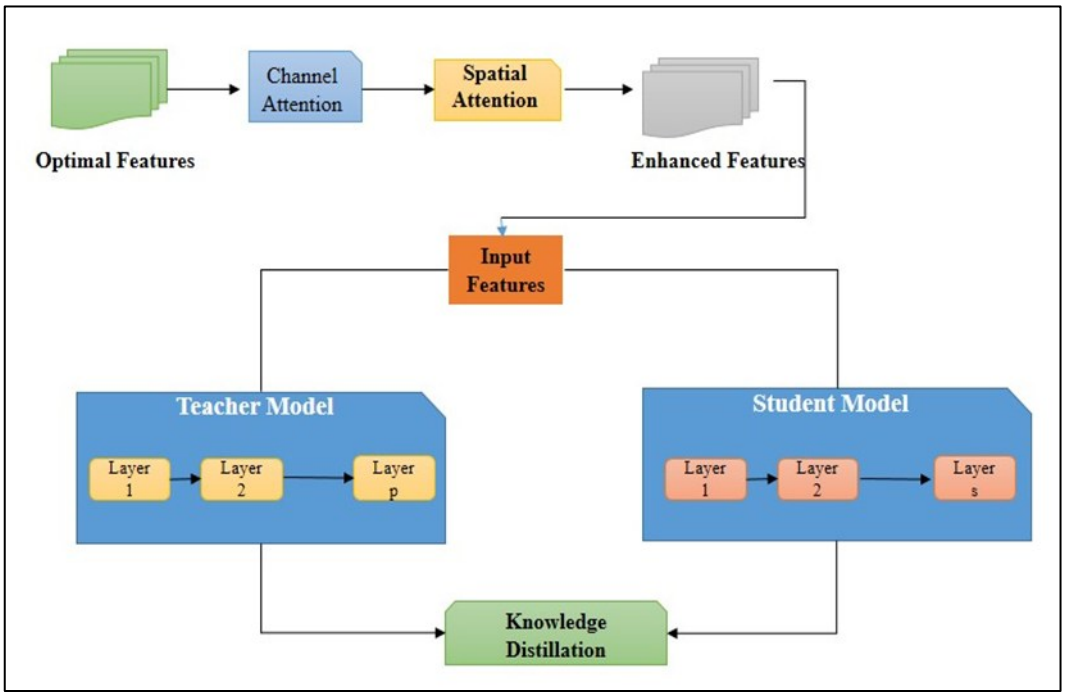


Figure 3. Architecture for Proposed GA-MSKAD Model

3.4.1 Teacher Module

In the teacher module, enhanced features are input into a transformer model that leverages self-attention to capture complex semantic relations among features, which may be irregular in the initial embeddings. The self-attention mechanism produces outputs as:

$$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{softmax}(\frac{QK^T}{\sqrt{d_x}})\text{V} \tag{4}$$

Where, d_x the feature dimension, and Q, K, and V is represent the query, key and value matrices derived from the initial embedding's combined with positional encoding:

$$Q = W_Q\, E \text{ , K= } W_k\, E, \text{ V= } W_V\, E \tag{5}$$

With W_Q, W_K, W_V as embedding weights. Multiple attention matrices are computed and integrated, followed by a layer normalization process LN (.) and a fully connected feed-forward network FFN (.).

$$Z = LN(Attention(Q, K, V) + E) \tag{6}$$

$$Output = LN(FFN(Z) + Z) \tag{7}$$

The final feature relation semantic encoding $F_{semantic}$ is obtained by applying mean pooling over the output:

$$F_{semantic} = MeanPooling(Output) \tag{8}$$

Where, $F_{semantic}$ represents the contextual semantic representation of the features.

3.4.2 Student Module

In the student module, a bidirectional gated recurrent unit with soft-attention (BiGRU-SA) captures global semantic features from the initial embedding's EEE, selectively emphasizing significant information. The soft-attention mechanism computes a weighted sum of feature-wise global embedding's:

$$F_{global} = \sum \alpha_i \, E_i \tag{9}$$

Where α_i is the attention weight or consideration score for each feature calculated as,

$$\alpha_i = \frac{\exp(W_a E_i)}{\sum_j \exp(W_a E_i)} \tag{10}$$

With W_a as a learnable weight matrix. The forward ($\underrightarrow{h_t}$) and backward ($\underleftarrow{h_t}$) GRU layers process the sequence to generate latent states:

$$\underrightarrow{h_t} = GRU_f(E_t, \underrightarrow{h_{t-1}}) \tag{11}$$

$$\underleftarrow{h_t} = GRU_f(E_t, \underleftarrow{h_{t+1}}) \tag{12}$$

Finally, the latent states from both directions are concatenated to form the global semantic encoding.

$$F_{final} = \left[\left[\underrightarrow{h_t}, \underleftarrow{h_t}\right]\right] \tag{13}$$

This approach aligns multi-level semantic knowledge through global attention and uses knowledge distillation to enhance classification accuracy in skin disease diagnosis.

3.5 Progressive Attention-based Multi-Scale Hierarchical Residual Swin Transformer

Progressive attention collects pre-processed data as input for the CNN and transformer, integrating both features and improving the latter features by channel attention. The features undergo global average pooling (GAP) after convolution, which is then activated by a sigmoid function while introducing a residual relation to enhance performance. Finally, the PAM model uses the convolution to accomplish the outcome features. The mathematical expression for the PAM evaluation process are represented in Eq. (14) and (15), respectively.

$$B_{si}^{l} = \lambda\left(BN\left(\vartheta_1\left(Concat\left(D_{si}^{l}, S_{si}^{l}\right)\right)\right)\right) \tag{14}$$

$$E_{si}^{l} = B_{si}^{l} * \rho\left(\vartheta_1\left(GAP\left(B_{ti}^{l}\right)\right)\right) + B_{si}^{l} \tag{15}$$

where D_{si}^{l} and S_{si}^{l} denote the features extracted by $l^{th}, (l = 2,3,4)$ a layer of CNN and transformer at time $si(i = 1,2)$, $Concat(\cdot)$ signifies the concatenation process, ϑ_1 indicates as 1×1 convolutional, terms as batch normalization, λ represents as ReLU activation process, $GAP(\cdot)$ refers to GAP, and then ρ indicates a sigmoid function and outcomes features of PAM referred as E_{si}^{l}.

3.6 Residual Lasso Logistic Regression Model

In this model, we use *M*, which denotes the total number of models of skin cancer through *Q* dimensional features. Therefore, the input feature *Y* represents the $M \times Q$ matrix, and the outcomes $Z = [x_1, x_2, \ldots, x_M]$ signify the binary variables. If the model is of target skin cancer; $x_j = 1$ or $x_j = 0$. Here, the logarithmic possibility of the j^{th} model fitting to the target skin cancer is expressed as

$$\log\left\{\frac{q\left(x_j = 1 \middle| Y^j, \alpha\right)}{1 - q\left(x_j = 1 \middle| Y^j, \alpha\right)}\right\} = \nu_\alpha\left(Y^j\right) \tag{16}$$

where $Y^j = [y_{j1}, y_{j2}, \ldots y_{jQ}]$ is a feature of the j^{th} sample, $\alpha = [\alpha_0, \alpha_1, \alpha_2, .. \alpha_Q]$ signifies the regression coefficient of LLRM, and $q\left(x_j = 1 \middle| Y^j, \alpha\right)$ terms as the possibility of the model, which fits the target skin cancer once regression coefficient are α.

4. Results and Discussion

In this section, a complete description of the dataset, a comparison of the experimental results with existing models, and a discussion are provided. The proposed model used 80% of the data for training and 20% for testing, as implemented in Python. The GA-MSKAD framework integrates multiple components: a transformer module with an attention dimension of 64, feed-forward network (FFN) layers of sizes 128 and 64, and an additional convolutional component consisting of 1D convolution layers with 256, 128, and 64 filters. For sequence modeling, the student model BIGRU-SA was employed, incorporating 64 GRU units with a dropout rate of 0.2. MaxPooling1D with a pool size of 2 and ReLU activation was applied to enhance feature extraction. The fully connected layers consisted of dense units (512, 256) with dropout rates of 0.3 and 0.2. Finally, the student model utilized zero-padding in 1D along with a Softmax activation function to generate multi-class classification outputs.

4.1 Dataset Description

In this section, the proposed model utilizes the unique HAM10000 dataset [29] [30] for segmentation, classification, and severity analysis. The dataset description is expressed as follows:

4.1.1 HAM10000 Dataset

Dermatoscopic images from several groups were obtained using various modalities. It contains 10,015 images with a resolution of 600 × 450 pixels, making it appropriate for academic machine learning training. The HAM10000 dataset consists of seven distinct skin lesion classes, each representing a different disease type. Melanoma is a serious and aggressive form of skin cancer originating from melanocytes, with 1,113 images. Melanocytic nevi, the most common category with 6,705 images, are benign moles formed by clusters of melanocytes. Basal cell carcinoma, represented by 514 images, is the most common skin

cancer, which grows slowly and rarely spreads. Actinic keratoses, with 327 images, are rough, scaly precancerous patches caused by sun damage. Benign keratosis-like lesions include seborrheic keratoses, solar lentigines, and lichen-planus like keratoses, with 1,099 images. Dermatofibroma, a harmless fibrous tissue growth appearing as firm nodules on the skin, is represented by 115 images. Finally, vascular lesions, such as angiomas, angiokeratomas, pyogenic granulomas, and hemorrhages, have 142 images. This diverse dataset provides a rich resource for training and evaluating deep learning models in skin cancer classification. Table 3 provides details of the dataset.

Table 3. Dataset Characteristics

Sr. No.	Parameter	Description
1	Dataset	HAM10000 (Human Against Machine with 10,000 training images)
2	Number of images	10,015 dermatoscopic images
3	Resolution	600 × 450 pixels (varies slightly for some images)
4	Skin lesion types	7 classes: Actinic Keratoses (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevi (NV), and Vascular Lesions (VASC)
5	Pre-processing techniques	Normalization, Contrast Enhancement, Resizing, Augmentation (Rotation, Flipping, Scaling, Cropping)
6	Data split	Commonly 80% Training, 20% Testing

4.2 Performance Evaluation

In this section, various graphical presentation evaluations are presented, including segmentation, classification, and severity analysis. In evaluation-level approaches for handling class imbalance, the use of balanced metrics rather than relying solely on raw accuracy can be misleading in imbalanced datasets. Metrics such as the F1-score, Precision-Recall AUC, balanced accuracy, and Matthews Correlation Coefficient (MCC) provide a fairer assessment of model performance across both benign and malignant categories. Additionally, applying stratified sampling when splitting the dataset into training, validation, and testing sets ensures that each subset maintains proportional representation of malignant and benign cases, leading to a more reliable and unbiased evaluation.

4.2.1 Experimental Analysis

The experimental results clearly highlight the impact of different batch sizes and learning rates on the performance of the proposed model. With a batch size of 8 and a learning rate of 0.01, the model achieved 96.38% accuracy. Lowering the learning rate to 0.001

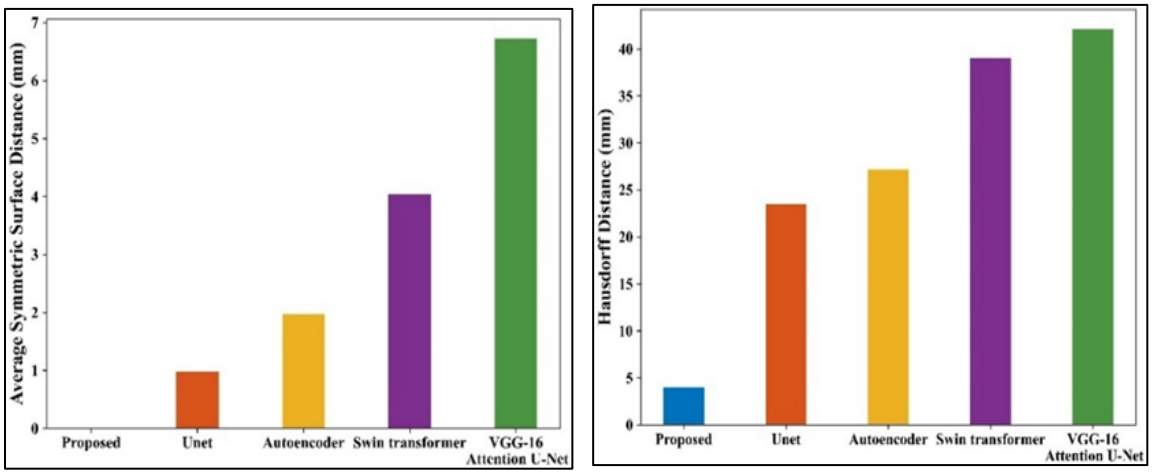
improved accuracy to 97.90%. Using batch size 16 with learning rate 0.01 gave 98.1% accuracy, while the best results of 98.9% accuracy and balanced metrics came with batch size 16, learning rate 0.001. These findings shown in Table 4 demonstrate that the choice of hyperparameters plays a crucial role in optimizing model performance, with smaller learning rates and moderate batch sizes leading to better generalization and overall accuracy.

Table 4. Evaluation of Hyper Parameter Configurations for Proposed Model

Sr. No.	Metrics / Methods	Proposed model			
		Accuracy	Precision	Recall	F1-score
1	Batch size = 8 & Learning rate = 0.01	96.38	96.53	97.10	96.89
2	Batch size = 8 & Learning rate = 0.001	97.90	96.89	97.43	97.01
3	Batch size = 16 & Learning rate = 0.01	98.1	97.34	96.23	97.87
4	Batch size = 16 & Learning rate = 0.001	98.9	98.83	98.1	98.91

4.2.2 Segmentation Analysis

This section explores the proposed PA-HRST segmentation model and compares it with existing models, namely U-Net [31], Autoencoder [32], Swin Transformer [33], and VGG-16 Attention U-Net [34].



(a) Average Symmetric Surface Distance (ASSD) (b) Hausdorff Distance (HD)

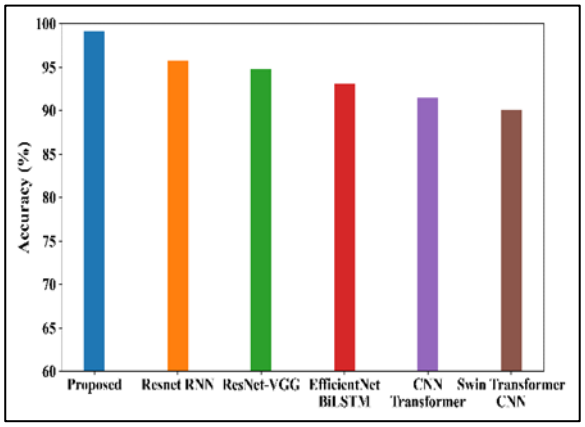
Figure 4. Segmentation analysis of (a) ASSD and (b) HD

Figure 4 shows the segmentation analysis of the proposed model compared with the current models. The HRST model collects enhanced features to facilitate semantic segmentation analysis. This model includes a residual connection to preserve relevant features

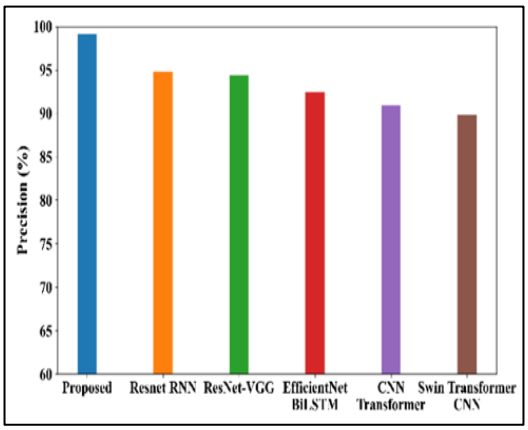
and an ST model to contribute to various features in different measures. The process is combined to produce enhanced segmentation performance. This model achieves an ASSD analysis of 0.008 and an HD analysis of 4%, which are compared with existing models to show their effective performance.

4.2.3 Classification Analysis

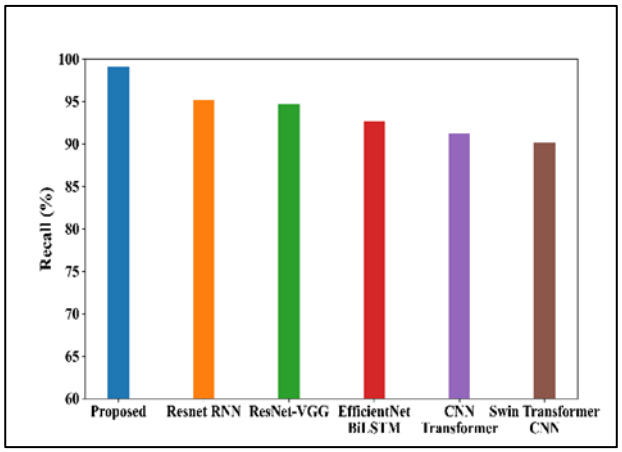
This section compares various existing techniques, including Swin-Transformer [33], ResNet-RNN [35], ResNet-VGG [36], EfficientNet-BiLSTM [37], and CNN-Transformer [38]. The performance of these models is evaluated based on key classification metrics: Accuracy, Precision, Recall, F1-score, and Specificity, with the results shown in Figure 5. These metrics offer a comprehensive view of each model's effectiveness, highlighting aspects such as overall performance, the ability to minimize false positives (Precision), the ability to detect true positives (Recall), the balance between Precision and Recall (F1-score), and the model's proficiency in correctly identifying negative cases (Specificity). However, the model obtains lower performance in malignant lesions that are harder to classify because of their visual similarity and class imbalances. Conversely, benign categories are rich in data, which makes them easier for the model to learn.



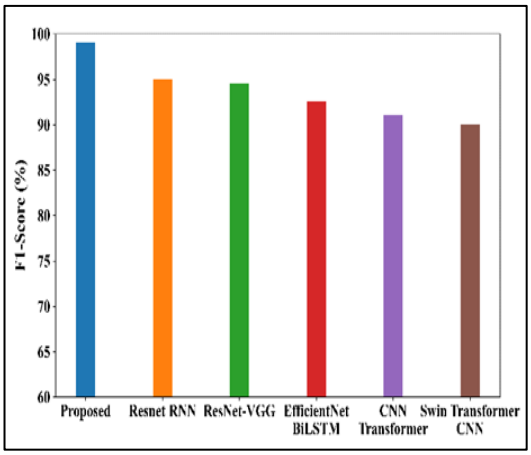
(a)



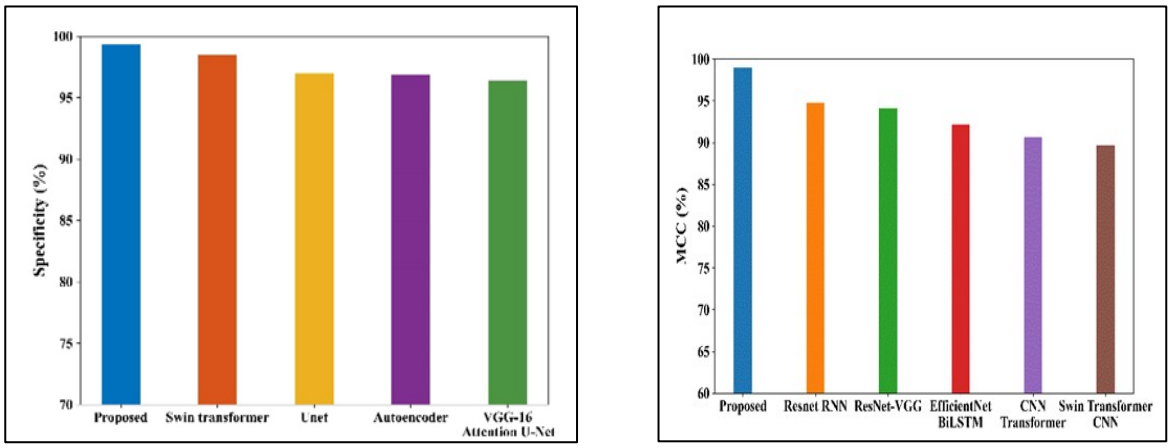
(b)



(c)



(d)



(e) (f)

Figure 5. Classification analysis of (a) Accuracy, (b) Precision, (c) Recall, (d) F1-score, (e) Specificity (f) MCC

The suggested model outperformed current methods including ResNet RNN, ResNet-VGG, EfficientNet BiLSTM, CNN-Transformer, and Swin Transformer CNN, achieving a high accuracy level of 99.18%. The proposed model outperformed the current ResNet RNN, which only managed a low accuracy level of 95.96%. The accuracies of the earlier models, ResNet-VGG and EfficientNet BiLSTM, were 94.73% and 93.06%, respectively. The existing fused CNN-transformer achieved 91.47% accuracy, while the popular Swin Transformer achieves 90.04% accuracy, which is less than the accuracy of the proposed and existing methods. The suggested model had a strong recall value of 99.13%. In addition, the precision of proposed model reached a high value of 99.09%, whereas the precision of the earlier models was lower and the proposed model achieved an MCC value of 98.96% demonstrating superior effectiveness in distinguishing between benign and malignant skin lesions.

The table below Table 5 highlights a comparison of the classification analysis between the proposed model and existing models.

Table 5. Comparison Analysis of the Proposed and Existing Models

Metrics / Models	Proposed model	Swin Transformer CNN [33]	ResNet-RNN [35]	ResNet-VGG [36]	EfficientNet-BiLSTM [37]	CNN-Transformer [38]
Accuracy (%)	99.18	90.04	95.76	94.73	93.06	91.47
Precision (%)	99.09	89.93	94.79	94.39	92.47	90.93
Recall (%)	99.13	90.13	95.13	94.65	92.7	91.17
Specificity (%)	99.03	89.92	95.01	94.47	92.53	91.16

F1-score (%)	99.11	89.98	94.96	94.52	92.59	91.05
MCC (%)	98.96	89.68	94.77	94.08	92.13	90.62

Figure 6 illustrates the analysis of the confusion matrix of the proposed model. Here, the ELPWF model reduces noise and enhances the quality of the input images. The proposed GA-MSKAD model includes global attention to capture and enhance features, and the MSKAD model classifies different types of Skin Cancer, namely AKIEC, BCC, BKL, DF, MEL, NV, and VASC. The proposed model correctly classified 680 labels but incorrectly classified 20 labels.

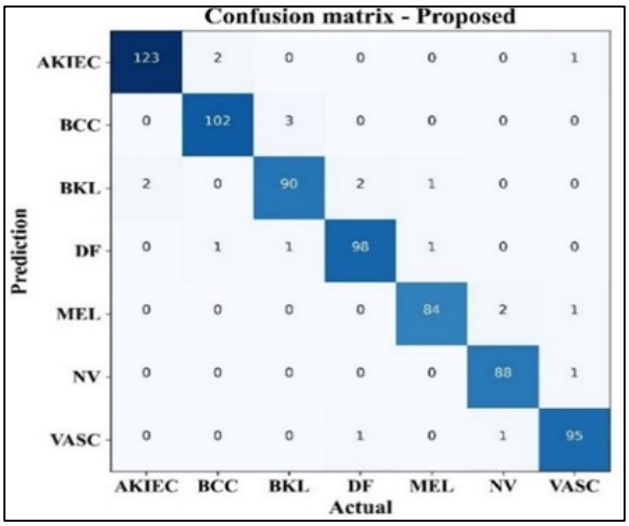
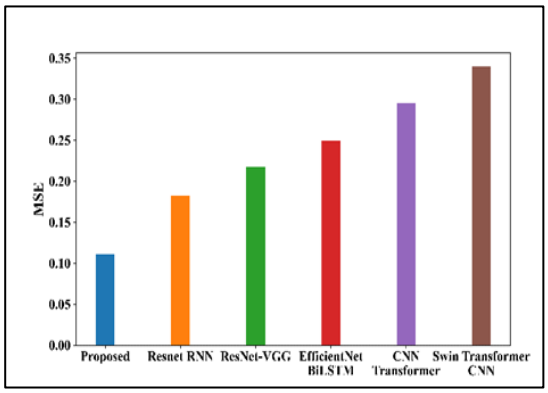
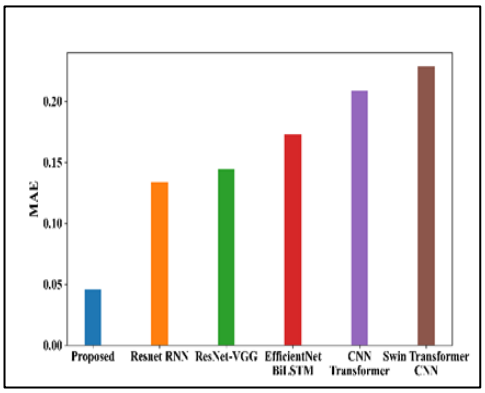


Figure 6. Confusion Matrix

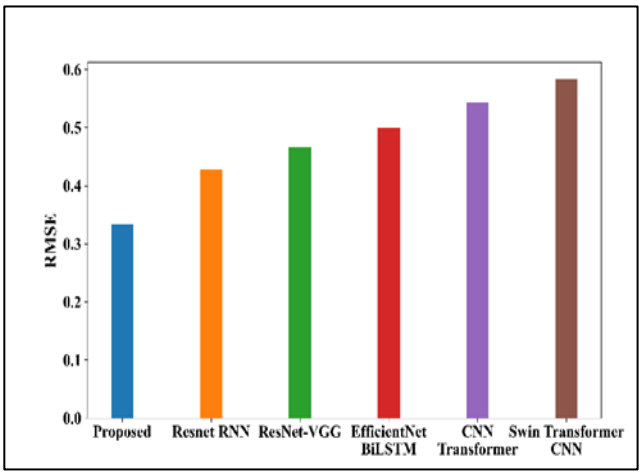
The comparison of the proposed model's MAE to current methods is shown in Figure 7. In this case, the proposed model outperforms earlier methods with a low MAE value of 4.57. The MAE values for the remaining methods, which included CNN-Transformer, ResNet-RNN, ResNet-VGG, and EfficientNet-BiLSTM, were 13.36, 14.43, 17.29, and 20.86, respectively. Additionally, the proposed model produces a low MSE score of 0.1114. The RMSE of the proposed model is then compared to previous methods, yielding a low RMSE of 33.38.



(a) Mean Squared Error (MSE)



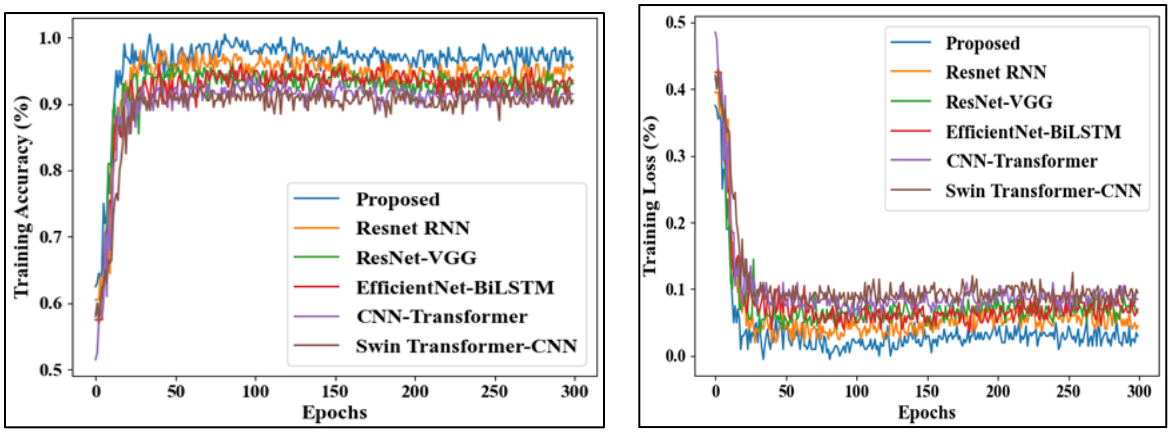
(b) Mean Absolute Error (MAE)



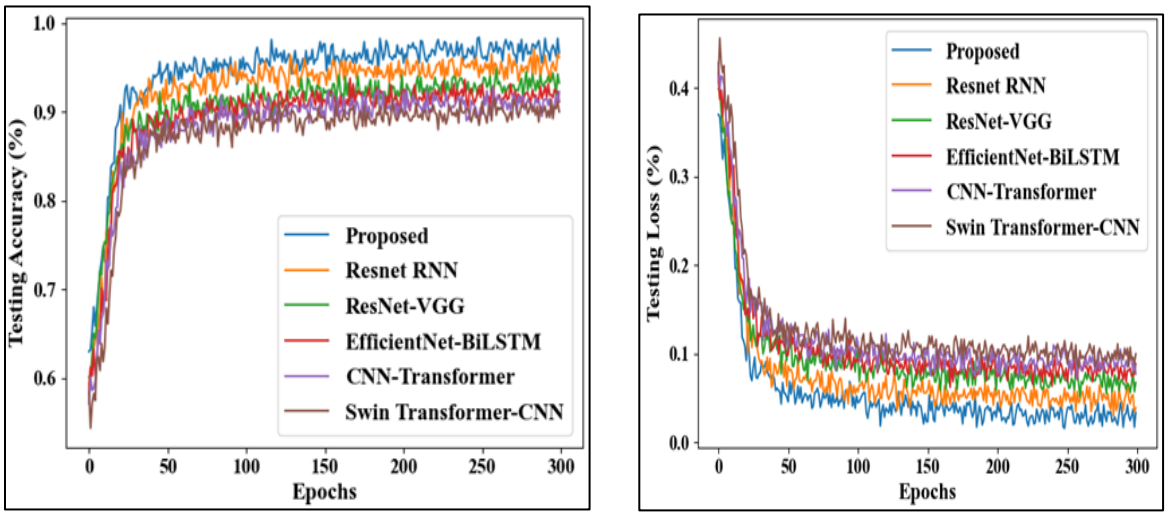
(c) Root Mean Squared Error (RMSE)

Figure 7. Analysis of (a) MSE, (b) MAE, and (c) RMSE

In experimental validation using real dermoscopic images, we enable simulation findings with benign or malignant classification trends that align with a controlled dataset. Here, performances show consistent analysis, demonstrating their flexibility and robustness in clinical applications. The figure below 8 provides a comprehensive view of the model's performance over time, displaying both training and testing accuracy, as well as the associated loss metrics. As the number of epochs progresses, the proposed model shows a steady improvement in both training and testing accuracy, indicating that it is effectively learning from the data and generalizing well to unseen samples. In parallel, the loss curves for both training and testing exhibit a consistent decline, suggesting that the model is converging toward an optimal solution with minimal error. The low loss values achieved during both training and testing phases further highlight the model's ability to make accurate predictions while avoiding overfitting, demonstrating its robustness and efficiency in handling the task at hand. This combined analysis of accuracy and loss underscores the model's effectiveness and stability across different stages of the training process.



(a) (b)



(c) (d)

Figure 8. Analysis of (a) Training Accuracy, (b) Training Loss, (c) Testing Accuracy and (d) Testing Loss

4.2.4 Severity Analysis

In this section, a suggested RLLM model is evaluated for severity analysis, as discussed below,

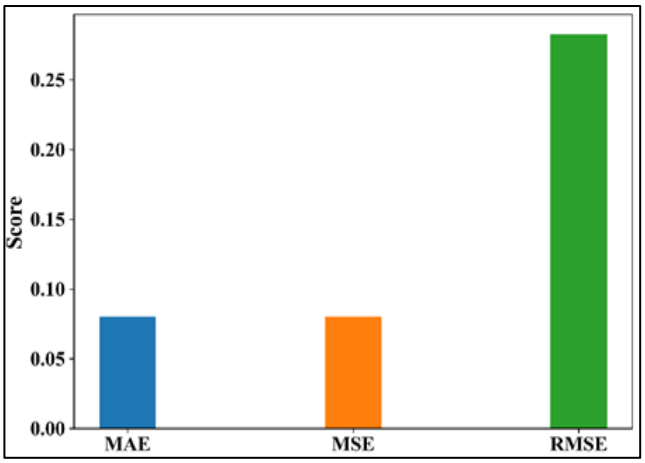


Figure 9. Severity Analysis of Proposed Model

Figure 9 displays the severity analysis of the proposed model. By identifying marginally significant characteristics, the suggested model standardizes the regression process using a Residual LASSO with logistic regression. Here, the LR model forecasts the severity of cancer while the residual LASSO model finds the most pertinent features consistent in noisy data. An RMSE of 0.282, an MAE of 0.08, and an MSE of 0.08 were achieved by this proposed model. Table 6 displays the analysis of the suggested model for severity.

Table 6. Proposed Model Analysis for Severity Analysis

Metrics	Proposed Model
MAE	0.08
MSE	0.08
RMSE	0.282

The proposed model provides segmentation, classification, and severity analysis, which are explored through performance analysis via graphical representation and tabulated in the overhead. This performance analysis is compared with existing models to demonstrate its high performance and enhance diagnostic accuracy. Table 7 provides a comparison between the proposed model and the current models.

Table 7. Comparison Analysis of the Proposed Model with Current Models

Methodology Used	Accuracy (%)
Capsule Network, SGD [16]	98.00
FrCN architecture [17]	97.98
GHO-Capsule Neural Network [18]	99.06
Tiny pyramid ViG [19]	95.52
CNN, DICE [20]	98.66
DCAN-Net [25]	97.57
Proposed Model	99.18

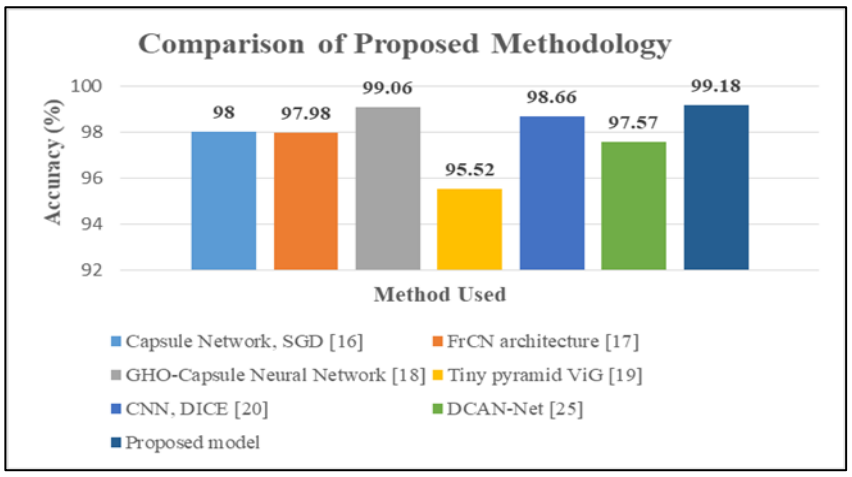


Figure 10. Comparison Chart of Accuracy between the Proposed Model and Existing Methods

Figure 10 shows Accuracy comparison of the proposed model with existing methodologies.

4.2.5 Ablation Study

The ablation study in Figure 11 illustrates the significance of each component in the proposed Severity-Aware Hybrid Deep Learning Framework for skin cancer imaging segmentation and classification.

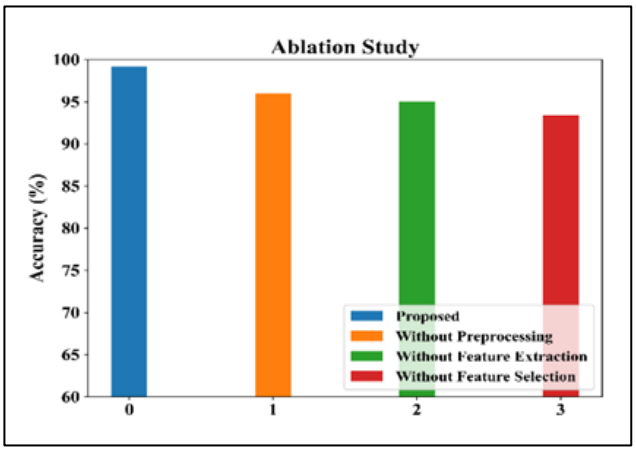


Figure 11. Ablation Study for Proposed Model

The efficacy of the integrated strategy was confirmed by the complete model, which had the best accuracy of 0.9918. The accuracy dropped to 0.96 when pre-processing was skipped, highlighting how crucial image enhancement and normalization are to improving model performance. Accuracy dropped to 0.9502 once the feature extraction module was removed, underscoring the critical function of TS-HCaps in capturing significant hierarchical lesion features. The accuracy dropped the most when feature selection was removed, reaching 0.9334. This highlights how crucial TCWOA-based feature optimization is for removing redundant data and choosing the most discriminative features. Overall, the ablation findings show that the high accuracy and robustness of the framework in identifying skin cancer lesions are achieved through the critical roles of pre-processing, feature extraction, and feature selection. Table 8 provides a detailed analysis.

Table 8. Ablation Study Analysis in Terms of Accuracy

Ablation Study	
Proposed model	0.9918
without-preprocessing	0.96
without-feature extraction	0.9502
without -feature selection	0.9334

5. Conclusion

The proposed method employs the HAM10000 dataset on Kaggle to develop a deep learning (DL)-based hybrid system for skin cancer segmentation, classification, and quantification of severity. The Enhanced Low-Pass Wiener Filter (ELPWF) is used for image quality enhancement and noise removal during preprocessing. Feature extraction from high-dimensional data is achieved by using the TS-HCaps model, effectively capturing temporal and hidden patterns. The Tent Chaotic Walrus Optimization Algorithm (TCWOA) then optimizes the features with reduced redundancy and computational costs. For segmentation, the PA-HRST model performs very well, with an ASSD of 0.008078 and an HD of 4%. Classification is achieved using a global attention-based hybrid model, enabling the GA-MSKAD framework to successfully differentiate seven skin cancer types (AKIEC, BCC, BKL, DF, MEL, NV, VASC) with 99.18% accuracy, 99.09% precision, 99.13% recall, and 99.11% F1-score. Finally, cancer severity is forecast through the RLLM regression model, and future work aims to incorporate Explainable AI (EAI) for enhanced interpretability.

References

[1] Aldhyani, Theyazn HH, Amit Verma, Mosleh Hmoud Al-Adhaileh, and Deepika Koundal. "Multi-class skin lesion classification using a lightweight dynamic kernel deep-learning-based convolutional neural network." Diagnostics 12, no. 9 (2022): 2048.

[2] Mohamed, Jalaleddin, Necmi Serkan Tezel, Javad Rahebi, and Raheleh Ghadami. "Melanoma Skin Cancer Recognition with a Convolutional Neural Network and Feature Dimensions Reduction with Aquila Optimizer." Diagnostics 15, no. 6 (2025): 761.

[3] Uthayakumar, G. S., Mallikarjun Yaramadhi, T. Marimuthu, and TR Vijaya Lakshmi. "Optimized mixed-order relation-aware recurrent neural networks based cad model for skin cancer detection and classification." Knowledge-Based Systems 315 (2025): 113222.

[4] Sol, Stefano, Fabiana Boncimino, Kristina Todorova, Sarah Elizabeth Waszyn, and Anna Mandinova. "Therapeutic approaches for Non-melanoma skin Cancer: Standard of Care and Emerging modalities." International journal of molecular sciences 25, no. 13 (2024): 7056.

[5] Akter, Mst Shapna, Hossain Shahriar, Sweta Sneha, and Alfredo Cuzzocrea. "Multi-class skin cancer classification architecture based on deep convolutional neural network." In 2022 IEEE International Conference on Big Data (Big Data), IEEE, (2022): 5404-5413.

[6] Tahir, Maryam, Ahmad Naeem, Hassaan Malik, Jawad Tanveer, Rizwan Ali Naqvi, and Seung-Won Lee. "DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images." Cancers 15, no. 7 (2023): 2179.

[7] Purni, JS Thanga, and R. Vedhapriyavadhana. "EOSA-Net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks." Journal of King Saud University-Computer and Information Sciences 36, no. 3 (2024): 102007.

[8] Veeramani, Nirmala, and Premaladha Jayaraman. "A promising AI based super resolution image reconstruction technique for early diagnosis of skin cancer." Scientific Reports 15, no. 1 (2025): 5084.

[9] Bala, Diponkor, Md Ibrahim Abdullah, Mohammad Alamgir Hossain, Mohammad Anwarul Islam, Md Atiqur Rahman, and Md Shamim Hossain. "Skinnet: An improved skin cancer classification system using convolutional neural network." In 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), IEEE, (2022): 1-6.

[10] Allugunti, Viswanatha Reddy. "A machine learning model for skin disease classification using convolution neural network." International Journal of Computing, Programming and Database Management 3, no. 1 (2022): 141-147.

[11] Ozdemir, Burhanettin, and Ishak Pacal. "A robust deep learning framework for multiclass skin cancer classification." Scientific Reports 15, no. 1 (2025): 4938.

[12] Aljohani, Khalil, and Turki Turki. "Automatic classification of melanoma skin cancer with deep convolutional neural networks." Ai 3, no. 2 (2022): 512-525.

[13] Yi, Jui-Yun, Sheng-Lung Huang, Shiun Li, Yu-You Yen, and Chun-Yeh Chen. "Discrimination of the Skin Cells from Cellular-Resolution Optical Coherence Tomography by Deep Learning." In Photonics, vol. 12, no. 3, p. 217. MDPI, 2025.

[14] Ali, Muhammad Danish, Tehseen Mazhar, Tariq Shahzad, Waheed Ur Rehman, Mohammad Shahid, and Habib Hamam. "An Advanced Deep Learning Framework for Skin Cancer Classification." The Review of Socionetwork Strategies 19, no. 1 (2025): 111-130.

[15] Alshahrani, Mohammed, Mohammed Al-Jabbar, Ebrahim Mohammed Senan, Ibrahim Abdulrab Ahmed, and Jamil Abdulhamid Mohammed Saif. "Analysis of dermoscopy images of multi-class for early detection of skin lesions by hybrid systems based on integrating features of CNN models." Plos one 19, no. 3 (2024): e0298305.

[16] Behara, Kavita, Ernest Bhero, and John Terhile Agee. "An improved skin lesion classification using a hybrid approach with active contour snake model and lightweight attention-guided capsule networks." Diagnostics 14, no. 6 (2024): 636.

[17] Adla, Devakishan, G. Venkata Rami Reddy, Padmalaya Nayak, and G. Karuna. "A full-resolution convolutional network with a dynamic graph cut algorithm for skin cancer classification and detection." Healthcare Analytics 3 (2023): 100154. https://doi.org/10.1016/j.health.2023.100154.

[18] Dubey, Vineet Kumar, and Vandana Dixit Kaushik. "Epidermis lesion detection via optimized distributed capsule neural network." Computers in Biology and Medicine 168 (2024): 107833.

[19] Santoso, Kevin Putra, Raden Venantius Hari Ginardi, Rangga Aldo Sastrowardoyo, and Fadhl Akmal Madany. "Leveraging spatial and semantic feature extraction for skin cancer diagnosis with capsule networks and graph neural networks." arXiv preprint arXiv:2403.12009 (2024).

[20] Salih, Omran, and Kevin Jan Duffy. "Optimization convolutional neural network for automatic skin lesion diagnosis using a genetic algorithm." Applied Sciences 13, no. 5 (2023): 3248.

[21] Sivasangeetha, A., G. Nallasivan, and M. Vargheese. "Skin cancer diagnosis using improved capsules network." DRRJ J 13, no. 5 (2023).

[22] Ali, Rizwan, A. Manikandan, and Jinghong Xu. "A novel framework of adaptive fuzzy-

GLCM segmentation and fuzzy with capsules network (F-CapsNet) classification." Neural Computing and Applications 35, no. 30 (2023): 22133-22149.

[23] Ahmed, Asaad, Guangmin Sun, Anas Bilal, Yu Li, and Shouki A. Ebad. "Precision and efficiency in skin cancer segmentation through a dual encoder deep learning model." Scientific Reports 15, no. 1 (2025): 4815.

[24] Chen, Yucong, Guang Yang, Xiaohua Dong, Junying Zeng, and Chuanbo Qin. "DSNET: A Lightweight Segmentation Model for Segmentation of Skin Cancer Lesion Regions." IEEE Access (2025).

[25] Thwin, Su Myat, Hyun-Seok Park, and Soo Hyun Seo. "A Trustworthy Framework for Skin Cancer Detection Using a CNN with a Modified Attention Mechanism." Applied Sciences 15, no. 3 (2025): 1067.

[26] Mohamed, Safia, Zaher Al Aghbari, and Ahmed M. Khedr. "Enhancing Skin Cancer Detection: An Integrated Approach Using Deep Learning and Metaheuristic Algorithms." In 2025 2nd International Conference on Advanced Innovations in Smart Cities (ICAISC), IEEE, (2025): 01-06.

[27] Saleh, Neven, Mohammed A. Hassan, and Ahmed M. Salaheldin. "Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making." Scientific Reports 14, no. 1 (2024): 17323.

[28] Majid, Abdul, Masad A. Alrasheedi, Abdulmajeed Atiah Alharbi, Jeza Allohibi, and Seung-Won Lee. "Modified whale optimization algorithm for multiclass skin cancer classification." Mathematics 13, no. 6 (2025): 929.

[29] https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification

[30] Serhani, Mohamed Adel, Asadullah Tariq, Tariq Qayyum, Ikbal Taleb, Irfanud Din, and Zouheir Trabelsi. "Meta-xpfl: An explainable and personalized federated meta-learning framework for privacy-aware iomt." IEEE Internet of Things Journal (2025).

[31] Hu, Binbin, Pan Zhou, Hongfang Yu, Yueyue Dai, Ming Wang, Shengbo Tan, and Ying Sun. "LeaNet: Lightweight U-shaped architecture for high-performance skin cancer image segmentation." Computers in Biology and Medicine 169 (2024): 107919.

[32] Reddy, Dasari Anantha, Swarup Roy, Sanjay Kumar, and Rakesh Tripathi. "A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification." Procedia Computer Science 218 (2023): 274-282.

[33] Pacal, Ishak, Melek Alaftekin, and Ferhat Devrim Zengul. "Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP." Journal of imaging informatics in medicine 37, no. 6 (2024): 3174-3192.

[34] Jimi, Anwar, Hind Abouche, Nabila Zrira, and Ibtissam Benmiloud. "SegSkin: An Effective Application for Skin Lesion Segmentation Using Attention-Based VGG-UNet." In Social Network Analysis and Mining Applications in Healthcare and Anomaly Detection, pp. 185-210. Cham: Springer Nature Switzerland, 2024.

[35] Zareen, Syeda Shamaila, Guangmin Sun, Mahwish Kundi, Syed Furqan Qadri, and Salman Qadri. "Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach." Computers, Materials & Continua 79, no. 1 (2024).

[36] Tabrizchi, Hamed, Sepideh Parvizpour, and Jafar Razmara. "An improved VGG model for skin cancer detection." Neural Processing Letters 55, no. 4 (2023): 3715-3732.
[37] Venugopal, Vipin, Navin Infant Raj, Malaya Kumar Nath, and Norton Stephen. "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images." Decision Analytics Journal 8 (2023): 100278.
[38] Pacal, Ishak, Burhanettin Ozdemir, Javanshir Zeynalov, Huseyn Gasimov, and Nurettin Pacal. "A novel CNN-ViT-based deep learning model for early skin cancer diagnosis." Biomedical Signal Processing and Control 104 (2025): 107627.