Leveraging Slicing Aided Hyper Inference-based Segmentation and Residual Gated Attention Vision Transformer for Multiclass Skin Cancer Classification

M Praveen Kumar
Dept. of ECE
SR University
Warangal, Telangana, India
praveenkumarjw@gmail.com

Sreedhar Kollem
Dept. of ECE
SR University
Warangal, Telangana, India
ksreedhar829@gmail.com

Samineni Peddakrishna
School of Electronics Engineering
VIT-AP University
Andhra Pradesh, India
krishna.samineni@gmail.com

***Abstract*- Skin cancer is a leading worldwide health issue, where accurate detection plays a significant part in increasing the intervention efficiency and patients' recovery values. Moreover, the intrinsic visual likenesses between the benign and malignant lesions also create an additional challenge in accurate lesion classifications. Early detection of skin cancer may help patients to treat the disease early and reduce the death rates. Although skin cancer detection is highly difficult because skin lesions often share a high degree of similarity. Hence, this work aims to develop an efficient multi-class skin cancer classification framework with an advanced segmentation and classification module. The dermoscopy images needed for performing multi-class skin classification are gathered through traditional public data sources. The collected images are directly fed into the Slicing Aided Hyper Inference (SAHI) network for performing skin lesion segmentation. This model helps to delineate the small lesion regions from the irregular shapes and borders. The segmented images are applied to the Residual Gated Attention Vision Transformer (RGA-ViT) for the classification of multi-class skin cancer, which helps to capture fine-grained local details, and hence the classification accuracy shows significant enhancement. The performance of the proposed segmentation and classification modules is analysed with the traditional works in terms of diverse measures to show their efficiency.**

Keywords-Multiclass Skin Cancer Classification; Slicing Aided Hyper Inference; Residual Vision Transformer; Gated Attention Mechanism

I. INTRODUCTION

Skin cancer represents the most frequently diagnosed malignancy across the globe, with both the melanoma and non-melanoma forms contributing significantly to illness and death [6]. According to the World Health Organization (WHO), non-melanoma skin cancers affect between 2-3 million people each year, while melanoma is about 1, 32,000 new cases each year [7]. So detecting the disease is significant, as its outcomes become much worse once it reaches its advanced stages. As per recordings, the 5-year survival rate for the early stage of melanoma is above 99%, but this decreases to less than 30% in metastatic cases [8]. Although early detection is crucial, it remains a significant challenge, particularly in regions with limited access to dermatologists and advanced diagnostic technologies.

Existing skin cancer detection methods depend on dermoscopy and clinical expertise, which often varies between observers, resulting in decreased consistency [9]. These limits highlight the high need for intelligent learning systems that could support clinicians in achieving expert-level accuracy for early detection. With the ongoing availability of the large-scale dermoscopic datasets, there is a strong opportunity to build data-driven frameworks capable of identifying subtle lesion features [10]. Nevertheless, existing automated systems continue to encounter several limitations, such as inadequate management of class imbalance and challenges in recognizing lesions with improper shapes, and restricted capacity [11]. Furthermore, many models lack clinically interpretable outputs. This underscores the need for a unified framework that can overcome these challenges by incorporating efficient modules to achieve reliable skin cancer classification.

Computer-aided diagnostic (CAD) approaches have been developed to distinguish various skin diseases from medical images with higher precision. Traditionally, these models are often relied on rule-based feature engineering [12], which was complex, lacked generalizability across diverse skin conditions, and was difficult to adapt for other diseases. The extracted features were then fed into learning models for identification. Early CAD approaches typically focused on a single disease, most often skin cancer, or on particular tasks [13]. To address the limitations of handcrafted features, deep learning methods were introduced. The success of these models largely depended on two factors: dataset size and the availability of computational resources for training [14]. A key objective of deep learning in dermatology is to design efficient diagnostic systems capable of handling the diverse characteristics of different skin diseases while delivering highly accurate results. However, many studies have remained biased, particularly skin cancer, due to the limited availability of public datasets. Moreover, most transfer learning approaches have relied on models pre-trained on ImageNet, which was not specifically designed for dermatological applications and therefore lacked generalization across all skin conditions [15]. To overcome these challenges, this work proposes a novel approach for skin cancer segmentation and classification.

The major contributions of the proposed model are illustrated in the following points:

- To develop a deep learning-based framework for accurate skin cancer classification, reducing human intervention and enabling early diagnosis to improve patient survival.
- To utilize the SAHI network for precise lesion segmentation, ensuring accurate localization of small, irregularly shaped, and malignant skin lesions.
- To implement an RGA-ViT model for multiclass skin cancer classification, combining Vision Transformer-based feature learning, residual connections for stable training,

and attention-guided feature selection for enhanced classification performance.

This study is organized into the following divisions. Division II presents the existing works along with their limitations and benefits. Division III describes the dataset information and the novel model. Division IV gives a clear explanation of the segmentation model, SAHI, along with the segmented results. Additionally, the working principle of the classification system, as RGA-ViT, is also stated in this section. Division V deals with the experimental investigation of the proposed system, supported by a set of graphs and tables. Finally, Division VI concludes the study and outlines future scopes.

II. Existing Works

A. Related Works

In 2025, Badr *et al.* [1] implemented a multi-model system for precise skin disease diagnosis. This framework used a five-class. The Xception model was employed to categorize skin lesions into five categories. Finally, the Bullous model reached higher values of accuracy, precision and recall.

In 2025, Sakib *et al.* [2] introduced a framework called LEVit that is designed to provide explainable and balanced classification of multiple skin lesion types. The model combines the strengths of two modality to learn the local features. Its performance was assessed using two benchmark datasets, where the convolutional layers helped capture fine lesion textures while the transformer component modelled global relationships across the image.

In 2025, AbuAlkebash *et al.* [3] proposed a model that includes a preprocessing stage to improve the accuracy of skin cancer diagnosis. It helps by detecting the region of interest, allowing the system to focus on the most important features. To reduce the impact of class imbalance in the dataset, several augmentation techniques were applied. Eight different pre-trained models were fine-tuned, and among them, the developed delivered the strongest performance.

In 2025, Prasanna *et al.* [4] produced a transformer-based model to enable fast and scalable classification of multiple

skin cancer types. The design brings together three important components: balanced image augmentation to address data distribution issues, hierarchical feature learning through masked patch reconstruction to capture deeper representations, and Error Level Analysis (ELA) during preprocessing to highlight subtle diagnostic details. Together, these elements strengthen the model's ability to detect fine-grained features while maintaining efficiency.

In 2026, Saeed *et al.* [5] introduced a TransXV2S-Net a specialized network for automated skin lesion classification. The framework is designed with multiple branches, each contributing to stronger feature representation. A key element of the architecture is the dual-path attention mechanism, which enhances discriminative learning by allowing the model to capture local textures while also preserving broader global structures. This combination helps the system achieve more accurate and reliable lesion identification.

B. Research Gaps and Challenges

Skin cancer classification faced several challenges that limited the efficiency and reliability of conventional approaches. These problems could be mitigated by developing advanced techniques, thereby improving model accuracy. Over the years, a wide range of approaches were proposed and tested, but many of these models continued to face significant issues. As a result, there was a strong demand for novel methods that could address the following limitations of existing approaches:

- Conventional models struggled with inefficient segmentation, which made it difficult to accurately identify the regions of the skin.
- Traditional classification models lacked effective global feature extraction and often encountered vanishing gradient problems.
- Earlier models did not incorporate powerful attention modules, leading to poor focus on intricate cancer-related patterns.
- Existing approaches typically relied on a single dataset to evaluate performance. Using multiple datasets would allow for better analysis of the system's ability to test with different cases.

III. Detailed View of Proposed Skin Cancer Classification with its Dermoscopy Images

A. Dataset Details

For the experimental validation, two publicly available benchmark datasets were utilized. The clear description of these datasets is provided below:

Dataset 1: The link for this is "https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic:": access date 2025-04-01. It consists of 9 diverse classes. It consists of thousands of labelled images across these 9 classes and provides high-quality dermoscopic images. Thereby, this dataset provides a diverse representation of skin cancer types, enabling efficient multiclass skin cancer classification.

Dataset 2: It is gathered from the link "https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection:": access date: 2025-04-01. It is widely used for binary and multiclass lesion analysis, and it offers a comprehensive foundation for testing the proposed frameworks. It consists of 3 different class labels.

The collected images from these datasets are represented as P_k . Here, the term $k = 1,2,...K$ denotes the total number of images. The reference images from these datasets are presented in Fig. 1.

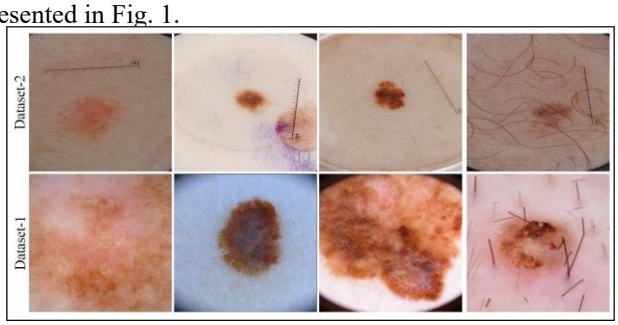


Fig. 1. Sample images collected from the datasets

B. Architectural Explanation of Novel Multiclass Skin Cancer Classification

Conventional models for skin cancer classification often face many issues, such as high computational complexity, inconsistent results, and slow manual inspection, which frequently lead to human error. To resolve these issues, a new

model is proposed and implemented in this work. It begins with the collection of relevant input images from trusted databases. After image acquisition, the collected images are subjected to a segmentation process, for which SAHI is employed to distinguish malignant or highly affected lesion areas from other regions. The resultant images that are segmented from SAHI are then transmitted to the classification network called RGA-ViT. In this network, a specially designed gated attention mechanism is introduced in the ViT to provide focused attention on subtle regions associated with disease prediction. Along with the attention mechanism, this network is also enhanced with residual connections in the ViT, which help capture a wide range of global dependencies and thus produce reliable classification results. To evaluate both SAHI and RGA-ViT, a set of experiments is conducted using previous models employed for skin cancer classification, along with a series of performance metrics. From these analyses, the outcomes of the proposed model are clearly validated, demonstrating that the developed framework contributes to the accurate recognition of skin cancer in the human body. The workflow of the new approach is provided in Fig. 2.

C. Segmentation System: SAHI

Objective: In skin cancer segmentation, one of the major difficulties lies in accurately detecting very small and irregular lesion regions, particularly those with fuzzy or unclear boundaries. These subtle details often occupy only a tiny portion of the dermoscopic image, which makes them hard for the model to capture during inference. To overcome this challenge, the SAHI method applies a slice-based detection strategy. By dividing the image into overlapping slices, small lesions appear larger within each slice, improving their visibility and making it easier for the model to identify them. This approach ensures that subtle diagnostic cues are not lost within the overall image and enhances the reliability of lesion segmentation.

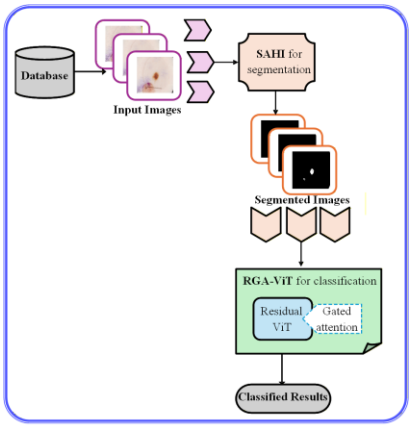
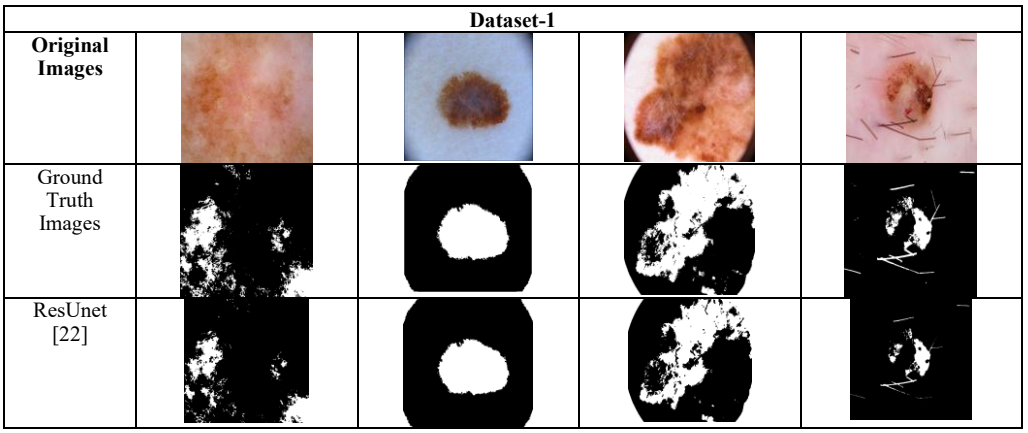


Fig. 2. Architectural view of the developed deep learning-based skin cancer classification system

Process flow of SAHI [16] [17]: The process begins with dermoscopic images P_k collected from public datasets, which are then fed into SAHI. Similar to the sliding window approach, SAHI partitions each image into multiple overlapping slices. This approach enlarges the appearance of small lesions within each slice, making them easier for the model to detect. The original image P_k is broken down into l overlapping patches q_l^l of a fixed size $M \times N$, as shown in Eq. (1).

$$I \rightarrow \{q_l^1, q_l^2, ...q_l^l\} \tag{1}$$

In this approach, each image is divided into patches $q_l^1, q_l^2, ...q_l^l$ that are reorganized while keeping their aspect ratio intact. These patches are then individually processed by SAHI to detect lesion regions, while the complete image is also examined to capture larger structural patterns. Once all patches have been analyzed, the results are combined. During this process, bounding boxes are compared. Boxes with values below a set threshold are removed, and for each lesion only the bounding box with the highest confidence score is preserved. This merging process ensures that small and irregular lesion regions are detected accurately without duplication. Finally, the obtained segmented images from this process are denoted as P_k^s. The resultant segmented images obtained from SAHI and other models are presented in Fig. 3 below.



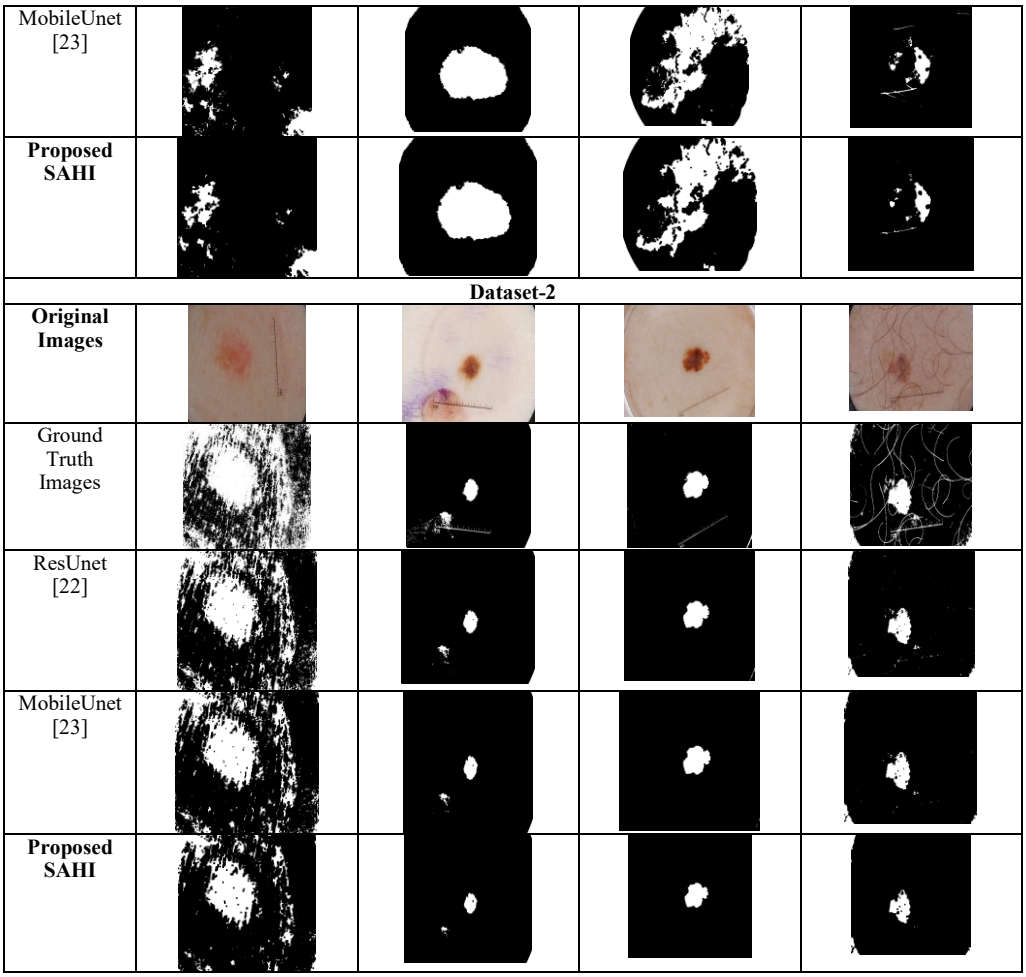


Fig. 3. Segmented images of the developed SAHI and other models

D. Vision Transformer

The ViT [18] is inspired by the original transformer model. Unlike traditional encoder-decoder structures, which compress a sequence into a single vector. ViT processes the input source and then portioned into very minute patches. Since transformers require one-dimensional embeddings, here each and every patches undergoes a linear projection to form a vector. This ensures that every data sample is arranged in a single row, forming a one-dimensional representation. After flattening into patches, the input to ViT is represented as Eq. (2).

$$y^l \in R^{N\times(P^2\times C)} \tag{2}$$

Here, the number of channels is C, and the patch size is P. Following the encoder layers, a trainable class token is introduced. Each patch is supplemented with a positional vector of equal dimension to retain spatial features. The ViT encoder applies several key operations. For each embedding, query q, key k, and value v vectors are computed. Attention scores are computed and concatenated across multiple heads. Residual connections add the input back to the attention output, stabilizing training. The process is summarized in Eq. (3) and Eq. (4).

$$z^l = M(LN(z^{l-1})) + z^{l-1} \tag{3}$$

$$z^l = M(LN(z^l)) + z^l \tag{4}$$

The transformed representation of the input sequence is z^l, layer normalization is LN, and the hidden state output from the $l-1$ encoder layer is z^{l-1}. Finally, the normalized class token output is defined in Eq. (5).

$$y \in LN(z_L^O) \tag{5}$$

Here, the class token embedding is represented as z_L^O, and it serves as the final representation for classification.

E. Classification System: RGA-ViT

Purpose: The RGA-ViT is the implemented model for skin cancer classification. It extends the standard ViT by integrating residual connections for stable learning and GA mechanisms to emphasizes clinically relevant lesion regions. This approach ensures comprehensive contextual representation and precise lesion detail analysis, supporting robust multi-class classification.

Process flow of the RGA-ViT is given below:

Input Layer: At first, the segmented dermoscopy image P_k^s from the SAHI model is input into the RGA-ViT. In this setup, the residual ViT [19] divides the images into fixed, distinct patches $P\times P$. Each patch is converted into a flattened vector of the given dimension d, as shown in Eq. (6).

$$x_i \in R^d,\ i = 1,2,..N \tag{6}$$

Here, the number of patches is N, and the patch embedding is x_i. Positional encoding is added to preserve spatial order, as given in Eq. (7).

$$x_i^{(O)} = Embed(x_i) + PE(i) \tag{7}$$

Each encoder block applies the patch embeddings, which are processed through a transformer encoder block. Mathematically, it is expressed in the Eq. (8) and Eq. (9).

$$b^{(l)} = b^{(l-1)} + M(b^{(l-1)}) \tag{8}$$

$$b^{(l)} = b^{(l)} + FFN(b^{(l)}) \tag{9}$$

Here, $b^{(l-1)}$ denotes the input to the encoder layer, input to the previous layer is $b^{(l)}$, and the feed-forward network is FFN.

GA [20]: To further improve classification, the GA module is integrated in the encoder part of the transformer. While the attention mechanism captures similarity between features by computing attention scores, the gated structure regulates the flow of this information, suppressing noise and emphasizing clinically relevant lesion regions. The gated output is defined in Eq. (10).

$$O = (U \otimes AV)W_o \tag{10}$$

The feature matrices derived from the input are a U, V learnable parameter matrix W_o. The attention weight matrix is A, which is computed in Eq. (11).

$$A = \mathrm{Re}\,LU(Q(Z)k(Z)^T + b) \tag{11}$$

The transformed output is Z $Q(\cdot)k(\cdot)$ a function similar to layer normalization with learnable parameters, and is a bias term. The input transformations are defined in Eq. (12).

$$Z = \varphi_c(XW_z), U = \varphi_u(XW_u), V = \varphi_v(XW_v) \tag{12}$$

The learnable weight matrices are W_z, W_u, W_v, and the activation function is $\varphi(\cdot)$. This GA mechanism ensures that patches with higher correlation to lesion characteristics are emphasized, while irrelevant information is suppressed. The gated embeddings are aggregated to form a global feature vector, as depicted in Eq. (13).

$$h = concat(x_1^{"}, x_2^{"}, ..x_n^{"}) \tag{13}$$

The gated embedding is $x_i^{"}$.

Residual connection: To stabilize training and preserve fine lesion details, residual connections are introduced in ViT. The residual structure follows the Residual Network) ResNet principle, as present in Eq. (14).

$$y = F(x, \{W_i\}) + x \tag{14}$$

The response of the residual block is y, while the input is x, and $F(x, \{W_i\})$ represents the transformation function.

Output Layer: This final representation is passed through a fully connected layer followed by softmax to give the accurate classification results across multiple skin cancer classes. The architecture of the RGA-ViT is shown in Fig. 4.

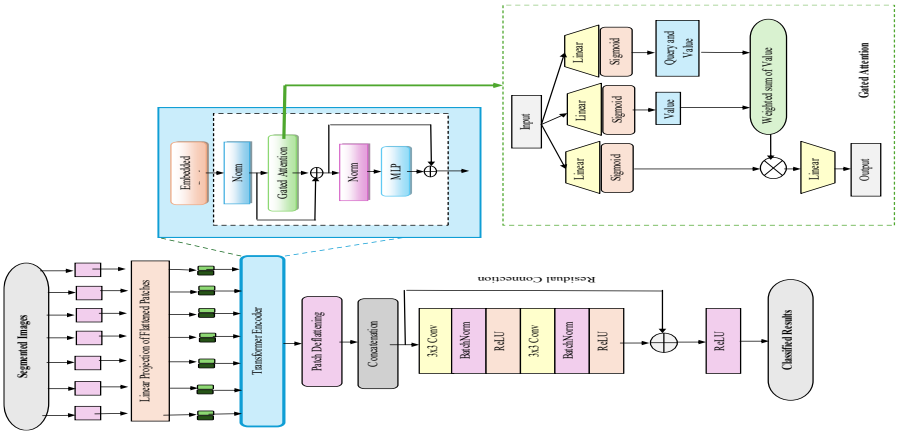


Fig. 4. Developed RGA-ViT for skin cancer classification

IV. Experimental Results Analysis

A. Simulation Setup

The developed model for skin cancer classification was executed on the Python platform. It was also analyzed with segmentation models such as Unet [21], ResUnet [22], MobileUnet [23], DenseUnet [24] and with the classification framework such as CNN [25], InceptionV3 [26], Resnet [27], VGG-19 [28], respectively. This comparison with the other models helps to understand the improvements of the developed system over the other approaches.

B. Assessment of the developed segmentation model on multiple frameworks

Fig. 5 depicts the SAHI results with other traditional approaches in terms of two datasets. This experimentation is carried out with certain performance measures with varying optimizers, to test how the model behaves under varying learning dynamics, because each optimizer has a unique characteristic that influences the model's performance. From Fig. 5 (a), the accuracy of the SAHI is superior to the other models, such as 13.2% of Unet, 10.2% of ResUnet, 8.1% of MobileUnet, and 3% of DenseUnet, particularly at the first dataset. Similarly, in the second dataset, the IoU value of the SAHI is greater than 17.8% of Unet, 21% of ResUnet, 15.7% of MobileUnet, and 7.3% of DenseUnet at the Adam optimizer. Hence, from this analysis, the predicted segmented regions overlap more closely with the ground truth and also reflect that a larger proportion of samples are correctly classified.

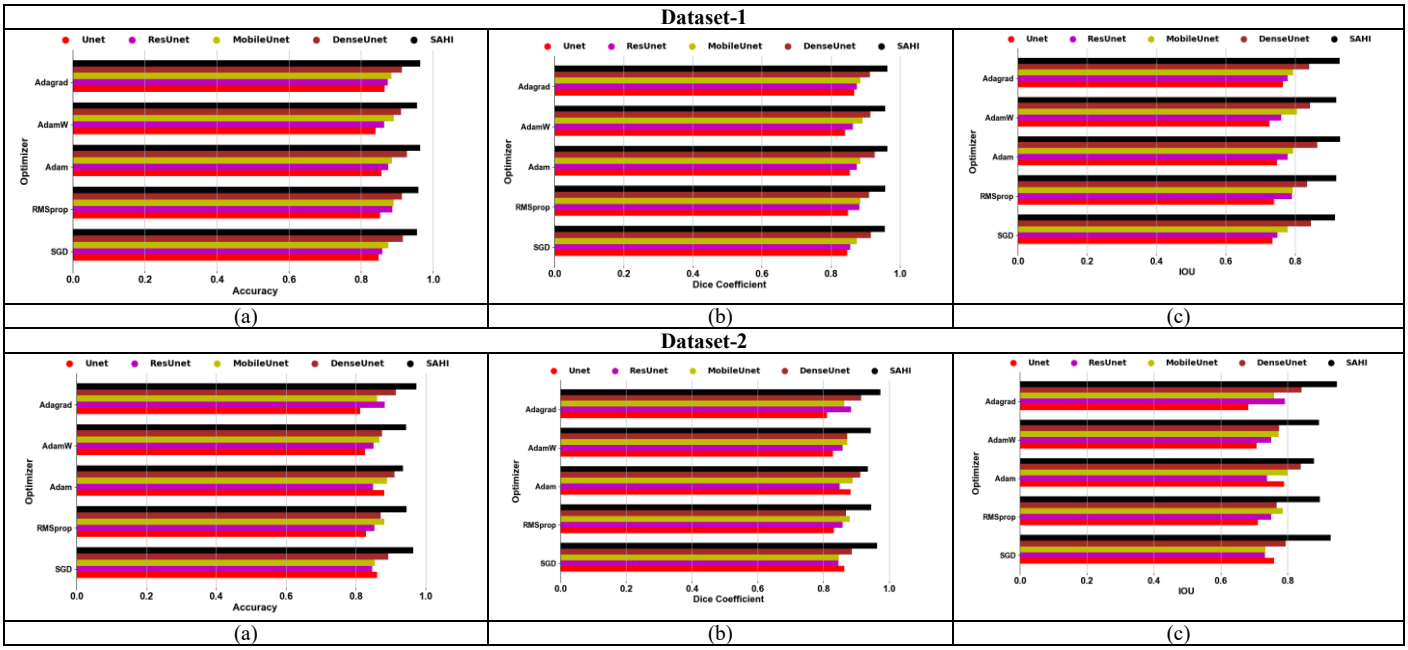
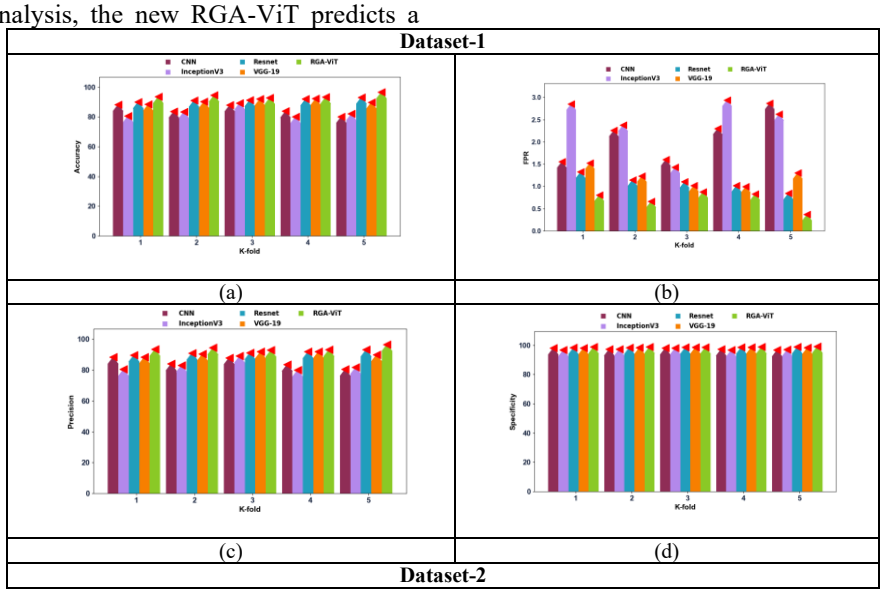


Fig. 5. Segmented results of the implemented SAHI versus prior models for (a) Accuracy, (b) Dice Coefficient, (c) IoU

C. *Validation of the classification performance of the RGA-ViT versus existing approaches*

Fig. 6 describes the classification analysis of the RGA-ViT with various k-fold rates. Testing with the numerous rates supports to identify whether the model is stable to input data distribution. In Fig. 6 (a), the RGA-ViT reaches a higher accuracy value, that is 7.7% of CNN, 6.6%$ of InceptionV3, 12.2% of Resnet, and 8% of VGG-19 at the second dataset. In Fig. 6 (c), the precision of the RGA-ViT also reached an optimal value compared with the CNN of 4.4%, InceptionV3 of 3.3%, and Resnet of 2.2%. and 1.1% of VGG-19 at the first dataset. Likewise, the RGA-ViT is also tested with the different batch values as given in Table II. In that table, the FPR of the RGA-ViT reaches a lower value than the other models, such as 47% of CNN, 71% of inceptionV3, 38% of Resnet, and 46% of VGG-19, especially in the second dataset. So overall from this analysis, the new RGA-ViT predicts a positive class more accurately, and a lower FPR implies that the model rarely misclassifies the negative samples. Therefore, the developed RGA-ViT is highly suitable for complex skin cancer classifications.



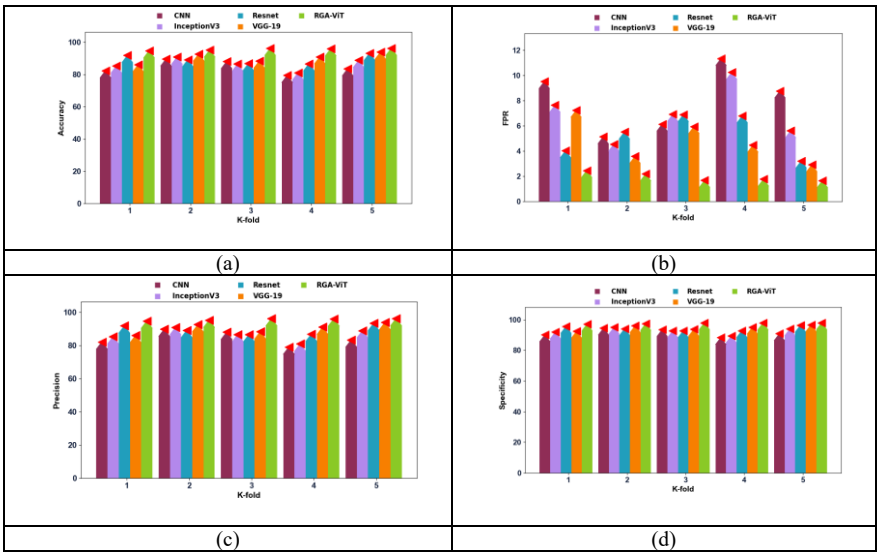


Fig. 6. Examination of RGA-ViT for skin cancer classification with the traditional baseline models with respect to (a) Accuracy, (b) FPR, (c) Precision, and (d) Specificity.

TABLE I. INVESTIGATION OF THE RGA-VIT WITH A CONVENTIONAL FRAMEWORK ACROSS TWO DIFFERENT DATASETS

Batch Sizes	CNN [25]	InceptionV3 [26]	Resnet [27]	VGG-19 [28]	RGA-ViT
Dataset-1					
Accuracy					
4	88.62019	81.07293	90.47282	88.74275	93.83868
8	84.10409	83.7741	91.38736	90.66139	94.85221
16	88.26663	89.70443	91.73149	92.36317	93.22114
32	84.18894	80.4271	92.32546	92.58002	93.62655
48	80.56381	82.3693	93.52284	90.21826	97.04898
FPR					
4	1.567103	2.864765	1.335492	1.536198	0.81834
8	2.268804	2.39066	1.158722	1.24805	0.677477
16	1.614143	1.444548	1.114795	1.034091	0.890944
32	2.313277	2.950211	1.028994	1.009185	0.842507
48	2.878526	2.632239	0.854605	1.317087	0.381056
FNR					
4	50.67233	65.15539	45.70991	50.36704	34.4306
8	60.16423	60.79602	42.98942	45.18689	30.26521
16	51.53488	47.86156	41.89734	39.80072	36.79218
32	60.05243	66.06408	39.93928	39.04895	35.25961
48	65.82437	63.15389	35.65801	46.45551	19.55681
Dataset-2					
Accuracy					
4	82.72727	85.72121	92.15758	86.47273	95.12727
8	90.03636	91.16364	89.58788	93.10303	95.58788
16	88.4	86.84848	87.15152	88.76364	96.61818
32	79.81818	81.47879	87.05455	91.38182	96.4
48	84	89.11515	93.55152	94.24242	96.66667
FPR					
4	9.549335	7.677242	4.069657	7.270955	2.497205
8	5.174396	4.587855	5.554491	3.629866	2.232143
16	6.17403	6.943632	6.926659	5.986526	1.719993
32	11.37966	10.28	6.836774	4.503135	1.815151
48	8.809618	5.647648	3.255551	2.958248	1.700869
FNR					
4	29.38064	25	14.56016	23.8141	9.292649
8	18.2058	16.28295	18.78508	12.81964	8.48991
16	20.76948	23.34294	22.70672	20.16367	6.542385
32	33.48554	31.2	23.01027	15.86876	6.980007
48	27.48741	19.75833	12.22752	10.89722	6.441394

V. Conclusion

In this study, a novel RGA-ViT framework was developed to overcome the limitations of conventional skin cancer classification methods. The framework begins with image acquisition, followed by lesion segmentation using the SAHI model, which divides images into smaller slices, processes them independently, and reconstructs the segmented output. The segmented images are then classified using the RGA-ViT model. By integrating residual connections and a GA-based attention mechanism, the framework effectively captures both global contextual information and fine lesion characteristics. Experimental results demonstrated superior performance over existing methods, achieving improvements of 7.1%, 4.4%, 7.2%, and 4.2% compared with CNN, InceptionV3, ResNet, and VGG-19, respectively.

References

[1] M. Badr, A. Elkasaby, M. Alrahmawy, and S. El-Metwally, "A multi-model deep learning architecture for Diagnosing multi-class skin diseases," J. Imaging Inform. Med., vol. 38, no. 3, pp. 1776–1795, 2025.

[2] A. H. Sakib, M. I. H. Siddiqui, S. Akter, A. A. Sakib, and M. R. Mahmud, "LEVit-Skin: A balanced and interpretable transformer-CNN model for multi-class skin cancer diagnosis," Int. J. Sci. Res. Arch., vol. 15, no. 1, pp. 1860–1873, 2025.

[3] H. AbuAlkebash, R. A. A. Saleh, and H. M. Ertunç, "Automated explainable deep learning framework for multiclass skin cancer detection and classification using hybrid YOLOv8 and vision transformer (ViT)," Biomed. Signal Process. Control, vol. 108, no. 107934, p. 107934, 2025.

[4] R. Jyothsna, K. Prasanna, U. Moulali, V. Surya Narayana Reddy, D. Ramya Krishna, and T. Praveen Kumar, "STViTDA-Net: An explainable transformer-based framework with STGAN-ViT-MAE and deformable attention for multi-class skin cancer classification," Intell. Based Med., vol. 13, no. 100350, p. 100350, 2026.

[5] A. Saeed, K. Shehzad, M. G. A. Malik, S. Ahmed, and A. T. Azar, "TransXV2S-NET: A novel hybrid deep learning architecture with dual-contextual graph attention for multi-class skin lesion classification," Knowl. Based Syst., vol. 337, no. 115407, p. 115407, 2026.

[6] M. I. Haque, K. Khatun, G. Rohman, A. Al Rakib, and M. M. Billah, "A deep learning framework for multi-class skin cancer diagnosis: Addressing class imbalance and performance optimization through Convolutional Neural Network architectures," Eur. J. Comp. Science Inform. Tech., vol. 13, no. 1, pp. 63–71, 2025.

[7] K. Muthulakshmi, S. Maruthuperumal, and G. R. Nesa Kumari, "An efficient multi-class dermatological lesion diagnosis using adaptive hybrid segmentation and residual graph CNN with attention mechanism," Biomed. Signal Process. Control, vol. 112, no. 108422, p. 108422, 2026.

[8] C. Srilakshmi, N. Ramakrishnaiah, and E. L. Lydia, "Semantic segmentation assisted deep ensemble feature learning model for skin-cancer detection and classification: SDENet," Complex Intell. Syst., vol. 12, no. 2, 2026.

[9] S. Gayen, S. Maity, S. Sarkar, E. Cuevas, and R. Sarkar, "SEFFNet: snapshot ensemble-based feature fusion network for skin cancer classification," Int. J. Mach. Learn. Cybern., vol. 17, no. 3, 2026.

[10] F. Bibi, Q. Abbas, and S. M. Naqi, "FoMoSkinNet: A dual-stream deep learning model for early detection and classification of non-dermoscopic skin lesions via focal modulation and local feature integration," IEEE Access, vol. 14, pp. 41998–42014, 2026.

[11] M. T. Noor, B. M. Shahria Alam, T. R. Orpa, S. A. Anika, M. T. Samiha, and F. Ahammed, "Deep learning for dermatology: An innovative framework for approaching precise skin cancer detection," arXiv [eess.IV], 2026.

[12] S. F. Fathima, P. R. K. Prasad, and V. Ramachandran, An efficient edge-based segmentation model with correlation weighted linked feature vector for melanoma detection using VGG16. In Smart Technologies and Intelligent Computing, 2026, pp. 346–352.

[13] A. U. Rahman, M. Albashrawi, B. Saqia, and Z. Halim, "Enhancing skin cancer detection using attention-based CapsNet with transfer learning from pre-trained CNN," Biomed. Signal Process. Control, vol. 116, no. 109544, p. 109544, 2026.

[14] S. Adablanu, U. Barman, D. Das, and T. J. Dweh, "Transforming skin cancer detection with AI-based convolutional and transformer models," iRadiology, vol. 4, no. 1, pp. 51–62, 2026.

[15] Mui-zzud-din, A. Naeem, H. Malik, M. Arsalan, A. Jafar, and U. Ali, "Skin cancer classification using a borderline-SMOTE enhanced neural network model on dermoscopic images," Biomed. Signal Process. Control, vol. 118, no. 109691, p. 109691, 2026.

[16] F. C. Akyon, S. O. Altinuc, and A. Temizel, "Slicing aided hyper inference and Fine-Tuning for small object detection," 2022 IEEE International Conference on Image Processing (ICIP), pp. 966–970, 2022.

[17] J. Lin, H. Lin, and F. Wang, "STPM_SAHI: A small-target forest fire detection model based on Swin Transformer and Slicing Aided Hyper Inference," Forests, vol. 13, no. 10, p. 1603, 2022.

[18] H. Shin, S. Jeon, Y. Seol, S. Kim, and D. Kang, "Vision transformer approach for classification of Alzheimer's disease using 18F-Florbetaben brain images," Appl. Sci., vol. 13, no. 6, p. 3453, 2023.

[19] Y. Liang et al., "Deep learning-based fluid identification with residual vision transformer network (ResViTNet)," Processes, vol. 13, no. 6, p. 1707, 2025.

[20] C. Zhao, J. Ye, Z. Zhu, and Y. Huang, "FLRNN-FGA: Fractional-order Lipschitz Recurrent Neural Network with frequency-domain Gated Attention mechanism for time series forecasting," Fractal Fract., vol. 8, no. 7, p. 433, 2024.

[21] W. Yin, D. Zhou, and R. Nie, "DI-UNet: dual-branch interactive U-Net for skin cancer image segmentation," J Cancer Res Clin Oncol., vol. 149, no. 17, pp. 15511–15524, 2023.

[22] S. Kahal, L. Meddeber, and T. Zouagui, Enhancing Skin Lesions Segmentation and Classification through Hierarchical Attention ResUnet Model. 2024 2nd Int. Conf. on Electrical Engineering and Automatic Control (ICEEAC), 2024, pp. 1–6.

[23] J. Deepa, V. Rekha, V. Kaliraj, S. B. Rani, and K. Aishwarya, "Automatic multi-scale dilated MobilenetV2 with attention-based lung nodule detection framework using adaptive 3D trans-MobileUnet++ segmentation," Sens. Imaging, vol. 26, no. 1, 2025.

[24] A. Al-Zubaidi, M. Al-Mukhtar, M. H. Al-hashimi, and H. Ijaz, "Skin lesion segmentation for melanoma using dilated DenseUNet," J. ICT Res. Appl., vol. 18, no. 1, pp. 21–35, 2024.

[25] J. Höhn et al., "Combining CNN-based histologic whole slide image analysis and patient data to improve skin cancer classification," Eur. J. Cancer, vol. 149, pp. 94–101, 2021.

[26] A. Demir, F. Yilmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3," 2019 Medical Technologies Congress (TIPTEKNO), pp. 1–4, 2019.

[27] A. Mehra, A. Bhati, A. Kumar, and R. Malhotra, "Skin cancer classification through transfer learning using REsNEt-50," in Advances in intelligent systems and computing, pp. 55–62, 2021.

[28] D. Shareen and N. Hossain, "Skin Cancer Detection using VGG-16," World J. Adv. Res. Rev., vol. 24, no. 2, pp. 2930–2937, 2024.