Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Shuffle Attention–Enhanced ResNet50 for Robust Skin Lesion Classification on Dermoscopic Images

Seyed Vahab Shojaedini[1*], Amir Nooraliei[2], Mahsa Monajemi[3]

1. Biomedical Engineering Department, Iranian Research Organization for Science & Technology, Tehran, Iran
2. Department of Computer and IT, Qa .c., Islamic Azad university , Qazvin , Iran
3. Department of Biomedical Engineering, Qa .c., Islamic Azad university , Qazvin , Iran

ARTICLE INFO

Article type:
Original Paper

Keywords:
Skin Lesion Classification
Shuffle Attention
Deep Learning
Dermoscopy

ABSTRACT

Introduction: Skin cancer is a prevalent global malignancy, and despite CNN advances in automated diagnosis, challenges in feature discrimination, robustness, and performance stability remain. Moreover, most previous studies primarily report single-run performance metrics, providing limited insight into the reproducibility and stability of deep learning models in clinically relevant settings.

Material and Methods: In this study, a Shuffle Attention–enhanced ResNet50 (SA-ResNet50) architecture was developed for automated skin lesion classification. The proposed model was systematically compared with baseline ResNet50 and ResNet50 integrated with the Convolutional Block Attention Module (CBAM). Experiments were conducted on two publicly available dermoscopic datasets, PH2 and HAM10000, using two data-partitioning strategies. To evaluate model robustness, independent-repeated ten times and K-fold cross validation methods were used for experiments, and performance variability was analyzed across repeated runs.

Results: The proposed SA-ResNet50 consistently outperformed the baseline models on both datasets while exhibiting lower performance variability. On the PH2 dataset, the proposed model achieved an accuracy of 98.9 ± 0.4%, sensitivity of 99.2 ± 0.3%, and specificity of 99.6 ± 0.2%. On the HAM10000 dataset with the first data partitioning strategy, SA-ResNet50 achieved an accuracy of 90.2 ± 1.6%, sensitivity of 85.7 ± 2.0%, and specificity of 92.8 ± 1.3%. In second strategy, the proposed model achieved an accuracy of 84.31±3.48%, Macro-F1 score of 71.86±7.23%, Balanced Accuracy of 70.29±9.36%, Matthews Correlation Coefficient (MCC) of 68.15±9.36, and an Area Under the ROC Curve (AUC) of 92.47±1.91, demonstrating robust performance on the highly imbalanced HAM10000 dataset with second data partitioning strategy.

Conclusion: The experimental results demonstrate that integrating the Shuffle Attention module into the ResNet50 backbone improves discriminative feature representation and enhances model robustness and convergence stability for dermoscopic skin lesion classification across benchmark datasets. These findings suggest that the proposed SA-ResNet50 constitutes an effective and reliable framework for automated skin lesion classification.

► *Please cite this article as:*
Shojaedini SV, Nooraliei A, Monajemi M. Shuffle Attention–Enhanced ResNet50 for Robust Skin Lesion Classification on Dermoscopic Images. Iran J Med Phys 2026; 23 (2): 132-141. 10.22038/ijmp.2026.96772.2718.

Introduction

Skin cancer is one of the most prevalent malignancies worldwide and continues to pose a serious public health challenge due to its rapidly increasing incidence and associated mortality rates(1). Among various skin cancer types, melanoma is the most aggressive and accounts for most skin cancer–related deaths despite its lower prevalence, owing to its high metastatic potential and rapid progression, making early and accurate diagnosis essential for improving patient survival and reducing treatment complexity(2),(3).

In conventional clinical practice, the diagnosis of skin lesions is primarily performed through visual inspection and dermoscopic examination by experienced dermatologists. Although dermoscopy improves diagnostic accuracy, expert-driven approaches remain subjective and highly dependent on clinician experience, leading to inter-observer variability and inconsistent diagnostic outcomes, particularly in visually similar benign and malignant lesions(4),(5).

These approaches relied on handcrafted features such as color, texture, and shape descriptors(6),(7), but their ability to generalize across diverse dermoscopic datasets and capture complex hierarchical image patterns remains limited (8).

In recent years, deep learning (DL), particularly convolutional neural networks (CNNs), has significantly advanced the field of medical image analysis. CNNs enable end-to-end feature learning directly from raw images without manual feature engineering, allowing effective extraction of both low- and high-level visual features and substantially improving dermoscopic image classification performance(9),(10) .

*Corresponding Author: Tel: +98-9123866360; Email: shojadini@irost.ir

Despite their success, conventional CNN architectures treat all spatial regions and feature channels uniformly, which may limit their ability to focus on diagnostically relevant regions. To overcome this limitation, attention mechanisms have been introduced to enhance feature representation by selectively emphasizing important spatial regions and informative feature channels. Several attention modules, including Squeeze-and-Excitation (SE)(11), Efficient Channel Attention (ECA)(12), Convolutional Block Attention Module (CBAM)(13), and Bottleneck Attention Module (BAM)(14), have been successfully integrated into CNN architectures.

However, although attention mechanisms have demonstrated promising results in dermoscopic image classification, different attention designs may provide varying benefits depending on their feature interaction strategies. For example, CBAM employs a sequential attention strategy in which channel attention is applied before spatial attention, which may limit the interaction between spatial and channel-wise features in complex lesion representations(13),(15). Recently, Shuffle Attention (SA) has been introduced as an efficient attention mechanism that models spatial and channel dependencies in parallel while enabling cross-group feature interaction through channel shuffling(16). Comprehensive evaluations comparing Shuffle Attention with established attention modules such as CBAM using the same backbone architecture remain limited.

Therefore, systematic comparisons between different attention mechanisms under identical backbone architectures are still necessary to better understand their relative effectiveness.

Another important limitation is the lack of robustness evaluation, as most studies report only single-run metrics such as accuracy or AUC without analyzing variability across multiple training runs (17),(18).

Motivated by these challenges, this study proposes a Shuffle Attention–enhanced ResNet50 framework for automated skin lesion classification. The proposed approach integrates SA into a residual learning backbone (ResNet50) to enhance feature discrimination while maintaining computational efficiency. The primary contribution of this work is not the introduction of a new attention mechanism, but the systematic evaluation of Shuffle Attention within a ResNet50 backbone through controlled comparison with CBAM and conventional ResNet50. In addition, this study investigates model robustness, stability and reproducibility by analyzing performance variability across multiple independent training runs.

Materials and Methods

In this study, a deep learning–based framework is proposed for automatic skin cancer classification using dermoscopic images. The primary objective of the proposed approach is to enhance the discriminative power of convolutional neural networks (CNNs) by explicitly modeling both spatial and channel-wise dependencies using an effective attention mechanism. As the backbone network, a ResNet-based architecture is employed due to its proven effectiveness in medical image analysis and ability to mitigate the vanishing gradient problem through residual connections(19).

Model Architecture

In the proposed SA-ResNet50 architecture, four Shuffle Attention blocks were integrated into the ResNet50 backbone. Each SA block was inserted at the end of one ResNet stage (Stage 1, Stage 2, Stage 3, and Stage 4), after the residual feature extraction process and before passing the refined feature maps to the next stage. This placement strategy enables progressive refinement of hierarchical features at different spatial resolutions while maintaining the original residual learning structure. For a fair comparison, the CBAM modules in the ResNet50+CBAM baseline were inserted at the same corresponding locations.

The overall framework of the proposed method is illustrated in Figure 1, which consists of three main components: The Stem as feature extraction using a ResNet backbone, four Stages each containing SA-Blocks, and the Classifier Head for 7 or 3-class classification.

The stem is responsible for extracting low-level visual features from the input image. Given an input image $I \in \mathbb{R}^{3\times H\times W}$,where H=W=224.

The network contains four stages, each operating at a different spatial resolution and channel depth. Each stage consists of a sequence of SA-Blocks. The number of channels in each stage is 256, 512, 1024, and 2048, respectively. After Stage 4, the feature map $X_4 \in \mathbb{R}^{2048\times 7\times 7}$ is passed to the classifier head, which is specifically designed for a 7-class (HAM10000 dataset) or 3-class (PH2 dataset) classification task. The Shuffle Attention–based architecture constitutes the main contribution of this study and is specifically designed to overcome the structural limitations of sequential attention mechanisms. The SA computes channel and spatial attention in parallel within grouped feature representations(16).

The architecture of the proposed Shuffle Attention module illustrates in Figure 2. As shown in the figure, the input feature map F is first divided into multiple channel groups (green arrow) along the channel dimension, following the Shuffle Attention formulation introduced in (16). This group-wise partitioning reduces computational complexity and enables the network to learn diverse feature representations within each group independently. SA first splits input feature map into G groups. After grouping, each sub-feature X_k is split into two parallel enhancement branches as X_{k1} and X_{k2}, namely the spectral (channel-wise) enhancement branch and the spatial enhancement branch, shown in equation 1 (orange boxes).

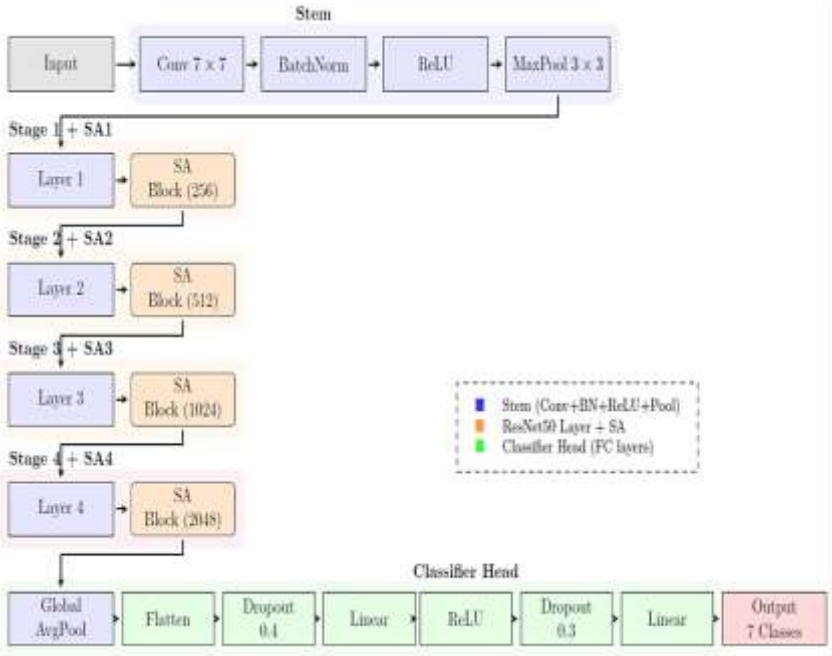


Figure 1. Architecture of the proposed method, Resnet50 layers and SA blocks

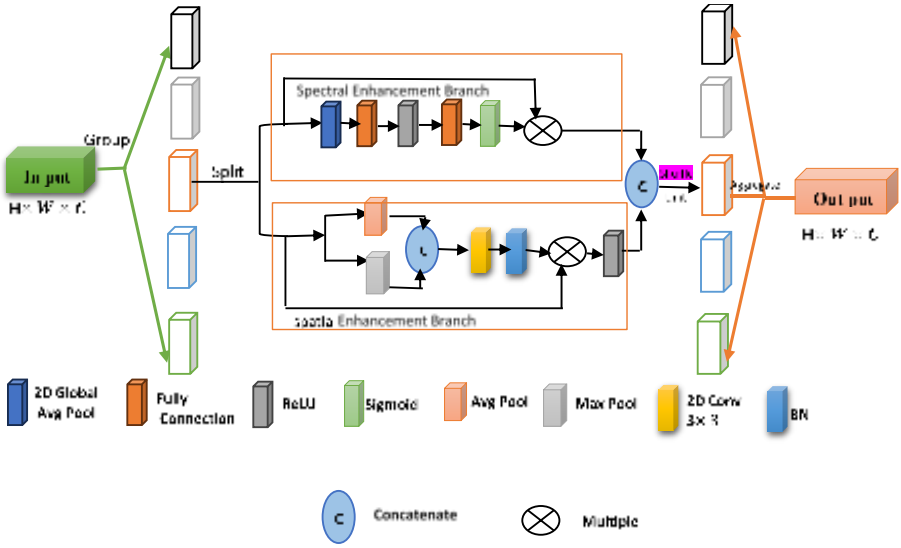


Figure 2. Architecture of the proposed Shuffle Attention module. The input feature map is first divided into channel groups, followed by parallel spectral (channel-wise) and spatial enhancement branches. The refined features are concatenated and redistributed through a shuffle unit to enable cross-group information interaction.

$$X \in \mathbb{R}^{C\times H\times W}$$

$$X = [X1, X2, \ldots . XG].\ \mathbb{R}^{\frac{C}{2G}\times H\times W}\ .k = 1 \ldots G \qquad (1)$$

$$X_k \rightarrow [X\_k1, X\_k2]].\ X_{k1}.X_{k2}\ \in \mathbb{R}^{\frac{C}{2G}\times H\times W}$$

In the spectral enhancement branch, global average pooling is applied to capture global contextual information across each channel group, followed by fully connected layers and nonlinear activations to generate channel attention weights, shown in equation 2 (16).

$$s = Fgp(X_{k1}) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{k1}(i.j) \in \mathbb{R}^{\frac{C}{2G}\times 1\times 1} \qquad (2)$$

After sigmoid activation block, the final output of channel attention can be obtained by equation 3.

$$X'_{k1} = \sigma(W_1 s + b_1)\cdot X_{k1} \qquad (3)$$

Where:

$W_1.\ b_1 \epsilon \mathbb{R}^{\frac{C}{2G}\times 1\times 1}$

σ is the sigmoid activation function

$(\sigma(z) = \frac{1}{(1+e^{-z})})$

The multiplication · is element-wise.

In parallel, the spatial enhancement branch employs average pooling and max pooling operations, together with convolutional layers, to learn discriminative spatial attention maps, emphasizing lesion-related regions. The output of spatial attention is obtained by equation 4.

$$X'_{k2} = \sigma\left(W_2 \cdot \hat{X}_{k2} + b_2\right) \cdot X_{k2} \quad (4)$$

Where:

$W_2 . b_2 \epsilon \mathbb{R}^{\frac{C}{2G} \times 1 \times 1}$

σ is the sigmoid activation function

$(\sigma(z) = \frac{1}{(1+e^{-z})})$

The multiplication · is element-wise.

The outputs of both branches are concatenated and passed through a shuffle unit, which redistributes feature channels across different groups to facilitate cross-group information interaction. Finally, the shuffled features are aggregated to produce an enhanced output feature map with the same spatial resolution and channel dimensionality as the input. The shuffle Attention structure illustrated in Figure 2. Also, the pseudocode of Shuffle attention shown in Figure 3.

Dataset

The proposed framework was evaluated on two publicly available dermoscopic image datasets, namely PH2 and HAM10000, which differ substantially in terms of dataset size, class distribution, and visual complexity. The main characteristics of the employed datasets are summarized in Table 1.

This study investigated model behavior under two distinct scenarios: a small-scale controlled dataset and a large-scale heterogeneous dataset(20),(21).

Data Preprocessing

The original HAM10000 images (600 × 450 pixels) were first resized shortest side to 256 while preserving aspect ratio then use random-resize-cropped with bilinear interpolation with scale (0.08, 1.0) to obtain a square region 224 × 224 pixels, which corresponds to the input size required by the network. This preprocessing pipeline avoids geometric distortion caused by direct resizing of rectangular images and helps preserve clinically relevant lesion characteristics, such as shape asymmetry. For the validation and test set, after resizing shortest side 256, used Center crop to reach 224 × 224.

ShuffleAttention

Input:
x - input features with shape [N, C, H, W]
cw, cb - channel attention parameters [1, C/2G, 1, 1]
sw, sb - spatial attention parameters [1, C/2G, 1, 1]
G - number of groups

Output:
out - output features with shape [N, C, H, W]

1: Get dimensions: N, C, H, W ← shape(x)

2: // Split channels into two groups
3: x_0, x_1 ← chunk(x, 2, dim=1)

4: // Channel Attention Branch
5: $\hat{x}_0$ ← GlobalAveragePooling(x_0) // Shape: [N, C/2, 1, 1]
6: $\hat{x}_0$ ← cw × $\hat{x}_0$ + cb // Scale and shift
7: $\hat{x}_0$ ← x_0 × Sigmoid($\hat{x}_0$) // Apply attention

8: // Spatial Attention Branch
9: $\hat{x}_1$ ← GroupNorm(x_1, G) // Group normalization
10: $\hat{x}_1$ ← sw × $\hat{x}_1$ + sb // Scale and shift
11: $\hat{x}_1$ ← x_1 × Sigmoid($\hat{x}_1$) // Apply attention

12: // Concatenate
13: out ← Concatenate([$\hat{x}_0$, $\hat{x}_1$], dim=1) // Shape: [N, C, H, W]

14: // Channel Shuffle
15: out ← channel_shuffle(out, 2)

16: return out

Figure 3. Pseudocode of Shuffle attention

Table 1. Detailed characteristics of the PH2 and HAM10000 datasets used in this study

Characteristic	PH2	HAM10000
Total images	200	10,015
Classes	3	7
Class distribution	CN: 80 AN: 80 MEL: 40	AKIEC: 327 BCC: 514 BKL: 1,099 DF: 115 MEL: 1,113 NV: 6,705 VASC: 142
Resolution	768 × 560	600 × 450
Image format	BMP	JPG
Imaging characteristics	Controlled dermoscopic images	Multi-source dermoscopic images
Class imbalance	Moderate	Severe (NV ≈ 67%)

Training Setup

All experiments were conducted using Python 3.9 and the PyTorch deep learning framework on a workstation equipped with an Intel Core i7-7700K CPU, 32 GB RAM, and an NVIDIA GeForce RTX 2060 GPU. The proposed SA-ResNet50 model was trained for 50 epochs with a batch size of 64 and an initial learning rate of 1e-4. The AdamW optimizer was employed with a weight decay of 1e-2. To prevent

overfitting, early stopping with a patience of 10 epochs was applied, and the model achieving the best validation performance was selected for final evaluation. The focal loss function was adopted to mitigate the effects of severe class imbalance in the HAM10000 dataset. The same training configuration was applied to all compared models, including ResNet50 and ResNet50+CBAM, to ensure a fair and reproducible comparison. The ResNet50 backbone was initialized using pretrained ImageNet weights.

Evaluation Protocol

To evaluate the model robustness and performance stability, we set two data partitioning strategies. In the first strategy, the dataset was randomly split as 70% training, 15% validation, and 15% testing, and each experiment was independently repeated ten times. The distributions of accuracy, sensitivity, and specificity obtained from these repeated runs were analyzed using box-and-whisker plots. In the second strategy, Patient-wise splitting and 5-fold was used because of data leakage problem in the HAM10000 dataset. Macro-F1 score, Balanced Accuracy, Matthews Correlation Coefficient (MCC), and AUC were additionally computed to provide a more comprehensive evaluation, particularly under the highly imbalanced HAM10000 dataset. These strategies enabled assessment of both predictive performance and convergence stability. All reported performance values are presented as mean ± standard deviation (SD) computed from the experimental runs.

Statistical Analysis

To further evaluate whether the observed performance differences between the evaluated models were statistically significant, statistical significance testing was performed using the results obtained from the ten independent experimental runs and a 5-fold cross-validation. The proposed SA-ResNet50 was compared with the two other models using a paired t-test. A significance level of $p < 0.05$ was adopted to determine statistical significance, and the corresponding p-values is reported in the Results section.

Results

We designed two data partitioning strategies. The first strategy (S1) is a random splitting, and the second strategy (S2) is stratified group-wise splitting. It is important to note that the performance of deep learning models on the HAM10000 dataset is highly dependent on the data-partitioning strategy. Although many existing studies report results obtained from image-level random or stratified splitting, such protocols may introduce lesion-level data leakage because multiple images of the same patient are present in the dataset(3),(9),(22),(23),(24),(25). Consequently, the training and testing sets may contain highly correlated samples, resulting in overly optimistic estimates of classification performance. To avoid this limitation, our experiments were conducted using StratifiedGroupKFold based on the patient identifier, ensuring that all images from a single patient were exclusively assigned to one fold. This evaluation protocol provides a more rigorous assessment of generalization ability, but we also set the strategy for numerical comparison with studies adopting image-level random splitting. Consequently, the reported results provide a more realistic assessment of model generalization.

First strategy

The dataset was divided randomly into training, validation, and test sets with a ratio of 70%, 15%, and 15%, respectively, for ten independent experimental runs.

Second strategy

The second partitioning strategy used to minimize the risk of patient-level data leakage, the HAM10000 dataset was partitioned using a stratified group-wise splitting strategy with five folds. Patient ID was used as the grouping variable, ensuring that all images belonging to the same patient were assigned to the same data partition and were not distributed across different partitions. This procedure prevents images from the same patient from appearing simultaneously in different partitions and thereby reduces the risk of data leakage and overly optimistic performance estimates. Stratification was applied to preserve the class distribution as closely as possible across the partitions. To ensure reproducibility, a fixed random seed (42) was used during data partitioning. Two strategies description show in Table 2.

Table 2. Data partitioning strategy used for the HAM10000 dataset

Strategy	Item	Description
S1	Dataset	HAM10000
	Splitting strategy	Stratified Group-wise
	Grouping variable	Patient ID
	Number of folds	5
	Stratification	Yes
S2	Train/Validation/Test	[70- 15- 15]
	Splitting strategy	Random split

To ensure a fair and controlled comparison, all experiments were conducted using ResNet50 as the backbone architecture for all evaluated models, including the baseline ResNet50, ResNet50 augmented with the Convolutional Block Attention Module (CBAM), and the proposed Shuffle Attention–based ResNet (SA-ResNet50).Except for the incorporated attention mechanisms, all network components, including convolutional layers, residual connections, classifier heads, and training configurations, were kept strictly identical across models. This controlled experimental design ensured that any observed differences in performance, robustness, or convergence behavior could be attributed solely to the employed attention strategies rather than architectural depth, parameter count, or training-related discrepancies.

Because the HAM10000 dataset exhibits a severe class imbalance, appropriate measures were adopted during training to reduce the dominance of the majority class. Specifically, focal loss was employed to improve the learning of minority lesion categories and to alleviate prediction bias toward melanocytic nevi.

The proposed SA-ResNet50 incorporates four Shuffle Attention blocks placed at the same network locations as CBAM to ensure a fair comparison setting. Unlike CBAM, Shuffle Attention jointly models spatial and channel-wise dependencies through feature grouping and channel shuffling, facilitating efficient cross-feature interaction with minimal computational overhead. models' computational complexity parameters show in Table 3**.**

Table 3. Comparison of the computational complexity and model size of the evaluated architectures.

Model	ResNet50	ResNet50 + CBAM	Proposed SA-ResNet50
Parameters (M)	24,560,711	25,257,423	25,258,447
Trainable Parameters	24,560,711	25,257,423	25,258,447
GFLOPs	4.133G	4.135G	4.135G
FPS	16.75	20.81	25.03

Table 3 compares the computational cost of the evaluated models. Although the proposed SA-ResNet50 incorporates the Shuffle Attention module, it introduces only a marginal increase in the number of trainable parameters and computational complexity while providing substantial improvements in classification performance.

Following the architectural design and controlled experimental setup, the performance of the evaluated models was assessed on the PH2 and HAM10000 datasets. In addition to conventional classification metrics, particular emphasis was placed on performance stability and robustness across multiple independent training runs. To this end, a box-and-whisker analysis was first conducted to examine the distribution and variability of accuracy, sensitivity, and specificity. Subsequently, a confusion matrix analysis was performed to provide detailed insights into class-wise diagnostic behavior. To evaluate robustness across repeated experiments, box-and-whisker plots of accuracy, sensitivity, and specificity obtained from ten independent runs on the PH2 dataset are presented in Figure 4.

This figure presents the distribution of accuracy, sensitivity, and specificity on the PH2 dataset, enabling a comprehensive assessment of both performance level and training stability. The baseline ResNet50 demonstrates strong and relatively stable performance across all evaluation metrics. The median accuracy stands at approximately 95.0±1.2%, with a sensitivity of 94.6±1.5% and a specificity of roughly 97.4±0.9%, reflecting stable convergence behavior. In contrast, the ResNet50+CBAM model shows markedly greater variability, achieving an accuracy of about 89.1±6.8%, a sensitivity of 89.0±8.7%, and a specificity of 94.1±6.2%, respectively.

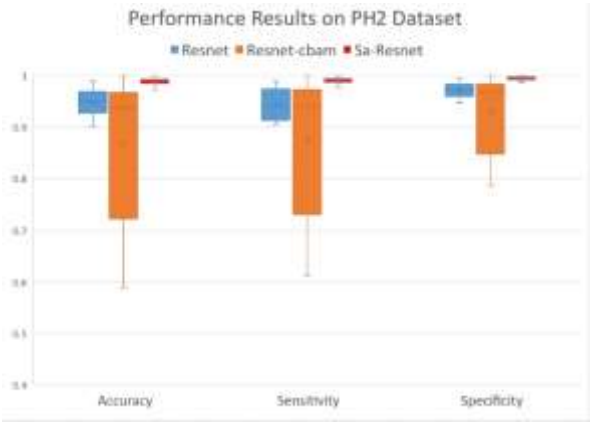


Figure 4. Box-and-whisker plots of accuracy, sensitivity, and specificity on the PH2 dataset.

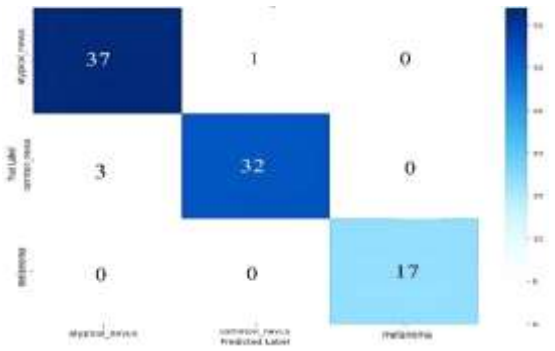


Figure 5. Confusion matrix of SA-ResNet50 on the PH2 dataset.

Although certain runs achieve competitive performance, the wide interquartile ranges and elongated lower whiskers reveal pronounced instability. In particular, accuracy and sensitivity show considerable dispersion, with some runs dropping significantly below the median. The proposed SA-ResNet50 achieves the highest and most consistent performance among all evaluated models. The median accuracy approaches 98.9±0.4%, sensitivity nears 99.2±0.3%, and specificity reaches approximately 99.6±0.2%, all characterized by exceptionally narrow interquartile ranges and minimal whisker lengths. These findings highlight the effectiveness of the Shuffle Attention mechanism in enhancing both discriminative capability and robustness under limited-data conditions. To further examine the diagnostic behavior of the proposed model, Figure 5 presents the confusion matrix of SA-ResNet50 on the PH2 dataset. The confusion matrix demonstrates near-perfect classification performance across all lesion categories with minimal misclassification. Importantly, melanoma samples are accurately detected with substantially reduced false-negative predictions. These findings further support the robustness and stability observed in the box-and-whisker analysis.

Following the analysis on the PH2 dataset, the evaluated models are further assessed on the HAM10000 dataset using two strategies to examine their robustness and

generalization capability under large-scale and highly imbalanced conditions**.** To analyze performance stability across 10 independent runs, box-and-whisker plots of accuracy, sensitivity, and specificity obtained from ten independent experiments on the HAM10000 dataset using first strategies are presented in Figure 6.

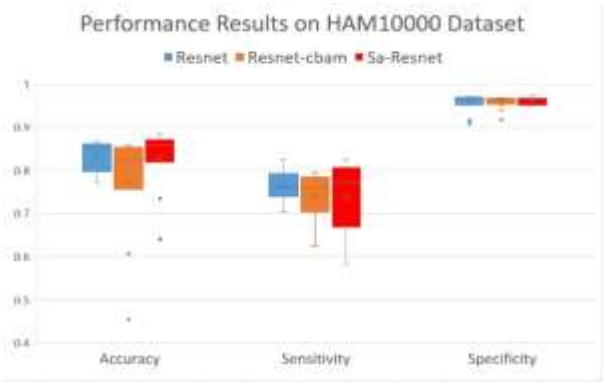


Figure 6. Box-and-whisker plots of accuracy, sensitivity, and specificity on the HAM10000 dataset (S1).

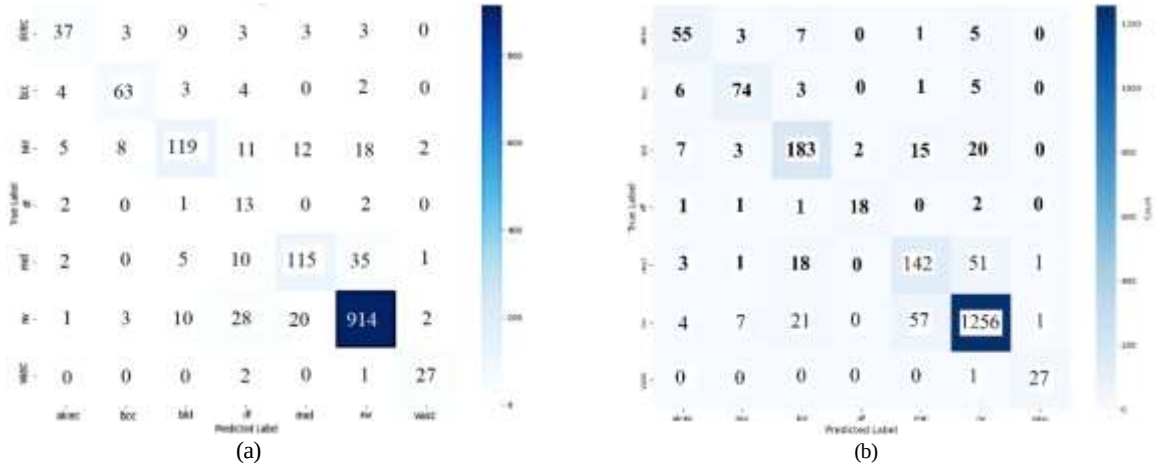
The baseline ResNet50 demonstrates stable yet moderately dispersed performance across the three evaluation metrics. The median accuracy, sensitivity, and specificity values are approximately 82.4±2.5%, 74.8±3.1%, and 88.9±1.8%, respectively**.** The ResNet50+CBAM model shows greater variability compared to the baseline. The performance metrics, namely accuracy, sensitivity, and specificity, approach values of 83.1±3.2%, 76.3±3.8%, and 89.5±2.1%, respectively. In contrast, the proposed SA-ResNet50 maintains comparatively tighter performance distributions. The model achieves approximately 90.2±1.6% accuracy, 85.7±2.0% sensitivity, and 92.8±1.3% specificity, demonstrating both improved performance and reduced variability across repeated experiments**.** These findings suggest that the Shuffle Attention mechanism enhances not only feature discrimination but also cross-class generalization**.** To provide a more reliable evaluation on the highly imbalanced HAM10000 dataset with the second data partitioning strategy, Macro-F1 score, Balanced Accuracy, Matthews Correlation Coefficient (MCC), and AUC were additionally computed. The corresponding results are presented in Table 4. The proposed SA-ResNet50 achieved the highest performance across these complementary evaluation metrics.

Table 4. Performance comparison of the evaluated models on the HAM10000 dataset with second data partitioning strategy using complementary evaluation metrics.

Model	ResNet50	ResNet50 + CBAM	Proposed SA-ResNet50
Accuracy	77.64±1.14	82.13±1.29	84.47±3.48
Bal. Acc. (%)	71.43±3.11	68.88±3.63	70.29±9.36
Macro-F1 (%)	67.25±2.67	68.38±3.91	71.76±7.23
MCC	61.89±1.30	66.11±2.21	68.15±9.36
Sensitivity	71.43±3.11	68.88±3.63	70.29±8.42
Specificity	95.47±0.29	95.62±0.16	95.79±1.00
ROC-AUC	90.36±1.50	90.58±2.69	92.47±1.91

To further investigate class-wise diagnostic behavior, Figure 7 presents the confusion matrix of SA-ResNet50 on the HAM10000 dataset. The confusion matrix indicates strong classification performance across both majority and minority lesion categories. High recognition rates are observed for melanoma and melanocytic nevi classes.

A limited number of misclassifications remain in underrepresented categories, which may be attributed to severe class imbalance and visual similarity among lesion categories within the HAM10000 dataset rather than limitations of the proposed architecture itself. In summary, the experimental results demonstrate that the proposed SA-ResNet 50 consistently outperforms both the baseline ResNet50 and ResNet50+CBAM across all evaluation metrics and datasets. On the small-scale PH2 dataset, SA-ResNet50 achieves near-perfect classification performance with minimal variability, highlighting its superior training stability and reliability under data-scarce conditions.



(a) (b)

Figure 7. Confusion matrix of SA-ResNet50 on the HAM10000 dataset, a) first strategy, b) second strategy.

On the large-scale, highly imbalanced HAM10000 dataset, SA-ResNet50 maintains higher median accuracy, sensitivity, and specificity compared to the other models, while demonstrating improved robustness and generalization across minority and majority classes. The confusion matrix analyses further confirm that SA-ResNet50 reduces critical misclassification errors, particularly false negatives in clinically important melanoma cases, and provides more balanced cross-class performance.

A statistical significance test based on ten experimental trials and a 5-fold cross-validation was carried out to ascertain the statistical significance of the observed performance differences among the examined models. Using a paired t-test, the SA-ResNet50 architecture suggested in this work was evaluated against the two remaining models. Table 5 displays the relevant p-values for the statistical significance level of $p < 0.05$.

Table 5. Statistical significance test for both datasets and two strategies

Ph2		
First strategy	P-value	t-statistic
ResNet vs ResNet-CBAM	0.1039	1.8078
ResNet vs SA-ResNet	0.0012	-4.6585
ResNet-CBAM vs SA-ResNet	0.0186	-2.8606
HAM10000		
Second strategy	P-value	t-statistic
ResNet vs ResNet-CBAM	0.245	1.31
ResNet vs SA-ResNet	0.010	4.07
ResNet-CBAM vs SA-ResNet	0.0004	5.83

Overall, these findings indicate that the Shuffle Attention mechanism effectively enhances feature discrimination, stability, and generalization, making SA-ResNet a promising approach for reliable, large-scale, multi-class dermoscopic image classification.

Discussion

The experimental evaluation demonstrates clear differences in performance, stability, and generalization among the three evaluated models across the PH2 and HAM10000 datasets.

The results obtained from the box-and-whisker and confusion matrix analyses reveal that the incorporation of attention mechanisms substantially influences predictive performance and model robustness under different experimental conditions. On the PH2 dataset, which is small and of high visual quality, all models achieve relatively high accuracy. However, SA-ResNet 50 consistently outperforms the other models across all metrics, with minimal variability across independent runs. This demonstrates not only higher performance but also remarkable training stability, which is particularly important when working with limited datasets where overfitting is a common concern. ResNet50+CBAM, while improving feature attention, shows greater variability, suggesting that sequential attention refinement may not fully stabilize learning on small-scale data. The baseline ResNet50 performs reasonably well but lacks the enhanced feature discrimination observed in SA-ResNet50, particularly in identifying melanoma cases.

On the more complex and highly imbalanced HAM10000 dataset, the differences among models become increasingly apparent, with SA-ResNet50 demonstrating stronger.

robustness and generalization capability than either ResNet50 or ResNet50+CBAM. The tighter performance distributions observed across repeated experiments suggest that Shuffle Attention contributes to more stable learning behavior under heterogeneous and imbalanced conditions compared with previous studies summarized in Table 6, the proposed SA-ResNet50 achieved competitive or superior performance on both the PH2 and HAM10000 datasets. Unlike several previous approaches that relied on extensive data augmentation or more complex architectures, the proposed method consistently demonstrated high classification performance together with low variability across experimental runs. Furthermore, compared with the sequential attention mechanism employed by CBAM, the Shuffle Attention module improved feature representation through parallel spatial and channel attention with channel shuffling, contributing to enhanced robustness and stability during training.

Sensitivity improvements indicate better detection of clinically significant melanoma cases, while specificity improvements reflect accurate classification of the majority classes.

In contrast, ResNet50 and ResNet50+CBAM exhibit broader distributions, particularly in sensitivity, reflecting difficulties in consistently identifying minority classes under highly imbalanced conditions.

Analysis of confusion matrices reinforces these observations. On PH2, SA-ResNet50 achieves near-perfect classification, correctly identifying all melanoma samples and minimizing false negatives. On HAM10000, SA-ResNet50 shows strong performance across both majority and minority classes, including melanoma (mel) and melanocytic nevi (nv), while reducing critical misclassifications in less-represented classes such as akiec and bcc.

Table 6. Comparison of the proposed SA-ResNet with previous studies on PH2 and HAM10000 dataset (S1, S2)

Study	strategy	Dataset	augmentation	Classes	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
(9)	S1	HAM10000	No	2	ResNet50	-	92.8	61.1
(22)		HAM10000	Yes	2	Advanced adaptive learning	87	91	84
(23)		HAM10000	Yes	2	CNN+sparrow search algorithm	94.75	-	-
(3)		HAM10000	Yes	7	ALBEF model	91.32	85.24	97.48
(24)		HAM10000	Yes	7	VGGNET-16	85.62	-	-
(26)		HAM10000	Yes	7	ResNet50	73.5	-	-
(25)		HAM10000	Yes	7	ResNet50	90	89	90
Proposed SA-ResNet		HAM10000	No	7	SA-ResNet50	90.2 ± 1.6	85.7 ± 2.0	92.8 ± 1.3
Proposed SA-ResNet	S2	HAM10000	No	7	SA-ResNet50	84.47±3.48	70.29±8.42	95.79±1.00
(7)	S1	PH2	No	3	ABCD rule/ SVM	92.5	100	87.5
(27)		PH2	No	3	ABCD rule/multi-thresholding	90.83	87.5	92.46
(28)		PH2	No	3	HAIM- EWHMLP	98.33	97.50	98.75
(29)		PH2	No	2	Fractal model	96.5	97.5	
Proposed SA-ResNet		PH2	Yes	3	SA-ResNet50	98.9 ± 0.4	99.2 ± 0.3	99.6 ± 0.2

The improved cross-class discrimination is attributed to the Shuffle Attention mechanism, which enhances interactions among feature groups and promotes stable learning across heterogeneous datasets**.**

The comparative results summarized in Table 5 clearly highlight the advantages of the proposed SA-ResNet50 architecture over both conventional ResNet50 and ResNet50+CBAM, as well as prior studies on PH2 and HAM10000 with two strategies. Across both datasets, SA-ResNet50 consistently achieves higher accuracy, sensitivity, and specificity, with minimal variability across independent runs. The Shuffle Attention mechanism contributes significantly to these improvements by enhancing cross-feature interactions, improving discrimination of minority classes, and stabilizing the learning process.

On PH2, the model demonstrates near-perfect performance even under data-scarce conditions, while on HAM10000, it effectively handles class imbalance and complex inter-class variability, providing robust and reliable predictions across all lesion categories. These findings confirm that the proposed approach not only outperforms existing methods in terms of raw metrics but also provides reproducible, clinically relevant results, making SA-ResNet50 a strong candidate for real-world automated skin lesion classification systems.

Limitations

Despite the promising performance of the proposed SA-ResNet50, several limitations should be acknowledged. First, the proposed model was evaluated only on the publicly available PH2 and HAM10000 datasets, and no external multi-center clinical dataset was available for independent validation. Second, although repeated experiments were performed to assess robustness, the study focused solely on image classification and did not investigate lesion segmentation or localization tasks. Finally, inference time and deployment on resource-constrained clinical devices were not evaluated. Future work will focus on external clinical validation, additional benchmark datasets, and real-world deployment studies.

Conclusion

In this study, a Shuffle Attention–enhanced ResNet50 (SA-ResNet50) framework was proposed for automated skin lesion classification using dermoscopic images. The proposed approach integrates the Shuffle Attention module into the ResNet50 backbone to improve feature representation by jointly modeling spatial and channel-wise dependencies**.**

The experimental evaluation on the PH2 and HAM10000 datasets using two strategies demonstrated that the proposed framework provides reliable classification performance and improved stability across repeated training runs compared with the baseline models. The robustness analysis and statistical evaluation further supported the consistency of the obtained results**.**

Overall, the findings indicate that SA-ResNet50 is a promising deep learning framework for dermoscopic skin lesion classification, particularly in challenging scenarios involving limited data availability and class imbalance. Future studies should focus on external validation using multi-center clinical datasets and further evaluation under real-world diagnostic conditions.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63.
2. Goyal M, Knackstedt T, Yan S, Hassanpour S.

Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. Comput Biol Med. 2020;127:104065.
3. Adebiyi A, Abdalnabi N, Hoffman E, Md S, Hirner J, Simoes EJ, et al. Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 Dataset. medRxiv. 2024;1–17.
4. Minarno AE, Fadhlan M, Munarko Y, Chandranegara DR. Classification of Dermoscopic Images Using CNN-SVM. JOIV Int J Informatics Vis. 2024;8(2):606–12.
5. Gurung S, Gao YR. Classification of melanoma (skin cancer) using convolutional neural network. In: 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA). IEEE; 2020. p. 1–8.
6. Nasir M, Attique Khan M, Sharif M, Lali IU, Saba T, Iqbal T. An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. Microsc Res Tech. 2018;81(6):528–43.
7. Zghal NS, Derbel N. Melanoma skin cancer detection based on image processing. Curr Med Imaging Rev. 2020;16(1):50–8.
8. Sreedhar B, BE MS, Kumar MS. A comparative study of melanoma skin cancer detection in traditional and current image processing techniques. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE; 2020. p. 654–8.
9. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. Eur J Cancer. 2019;111:148–54.
10. Vivek N, Ramesh K. Deep Learning Multi-Modal Melanoma Detection: Algorithm Development and Validation. JMIR AI. 2025;4:e66561.
11. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 7132–41.
12. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 11534–42.
13. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2018. p. 3–19.
14. Park J, Woo S, Lee JY, Kweon IS. Bam: Bottleneck attention module. arXiv Prepr arXiv180706514. 2018;
15. Ding X, Qian K, Zhang Q, Jiang X, Dong L. Dual-channel compression mapping network with fused attention mechanism for medical image segmentation. Sci Rep. 2025;15(1):8906.
16. Zhang QL, Yang YB. Sa-net: Shuffle attention for deep convolutional neural networks. In: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2021. p. 2235–9.
17. Nasir IM, Alshaya H, Tehsin S, Bouchelligua W. Transformer-Based Foundation Learning for Robust and Data-Efficient Skin Disease Imaging. Diagnostics 2026, Vol 16, Page 440. 2026 Feb 1;16(3):440.
18. Hark BD. Beyond Accuracy: Methodological Evaluation of Performance Stability in Deep Learning Models on the HAM10000 Multi-Class Skin Lesion Benchmark. Türkiye Klin Biyoistatistik. 2025;17(3):140–50.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770–8.
20. Mendonça T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J. PH 2-A dermoscopic image database for research and benchmarking. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2013. p. 5437–40.
21. Tschandl P, Rosendahl C, Kittler H. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5.
22. Khan MA, Mazhar T, Ali MD, Khattak UF, Shahzad T, Saeed MM, et al. Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks. Discov Oncol. 2025;16(1):645.
23. Balaha HM, Hassan AES. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Comput Appl. 2023;35(1):815–53.
24. Çevik E, Zengin K. Classification of skin lesions in dermatoscopic images with deep convolution network. Avrupa Bilim ve Teknol Derg. 2019;309–18.
25. Kondaveeti HK, Edupuganti P. Skin cancer classification using transfer learning. In: 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI). IEEE; 2020. p. 1–4.
26. Reddy ND. Classification of dermoscopy images using deep learning. arXiv Prepr arXiv180801607. 2018;
27. Maniraj SP, Maran PS. A hybrid deep learning approach for skin cancer diagnosis using subband fusion of 3D wavelets. J Supercomput. 2022;78(10):12394–409.
28. Vidya Lakshmi V, Leena Jasmine JS. A Hybrid Artificial Intelligence Model for Skin Cancer Diagnosis. Comput Syst Sci Eng. 2021;37(2):233–45.
29. Jacob S, Rosita JD. Fractal Model for Skin Cancer Diagnosis Using Probabilistic Classifiers. Int J Adv Signal Image Sci. 2021;7(1):21–9.