



ARTICLE

Hybrid Attention-Enhanced Transfer Learning Framework with Particle Swarm Optimization for Explainable Skin Cancer Classification

Chanchal Biswas¹, Naveenkumar Mahamkali¹, Satish Anamalamudi¹, Bhaskar Marapelli² and Ch Anil Carie^{1,*}

¹Department of Computer Science and Engineering, SRM University AP, Amaravati, Andhra Pradesh, India

²Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

*Corresponding Author: Ch Anil Carie. Email: anilcarie.c@srmap.edu.in

Received: 16 June 2026; Accepted: 28 August 2026; Published: 21 September 2026

ABSTRACT: Effective feature extraction is problematic due to subtle differences in skin texture, color, and shape. Moreover, imbalanced datasets are common in medical image analysis, which complicates classification by biasing models toward dominant classes, causing overfitting. The proposed approach is a hybrid one comprising transfer learning and self-attention. We have employed ResNet-50 pre-trained model as a feature extractor for dermatoscopic images. Max Pooling and Global Average Pooling were used to focus on the important patterns and reduce irrelevant background information. The normal kernel and L2-regularization initialization are applied to emphasize salient patterns and suppress irrelevant background information. The module of self-attention comes after the max-pooling and before the global average-pooling. This method allows for efficient modeling of lesion relationships across the entire globe in a dual pooling strategy. We have used particle swarm optimization (PSO) to tune the hyperparameters. Optimization methods, such as early stopping, ReduceLROnPlateau, dropout, and 10-fold cross-validation, help to guarantee strong convergence and thus avoid overfitting. In addition, we have used data augmentation and transfer learning to overcome the imbalance in the datasets. This will help to generalize across different lesion types. The proposed model achieves an accuracy of 97.77%, precision of 98.00%, recall of 97.17%, and F1-score of 97.45% on the HAM10000 dataset. To assess the robustness and generalization ability of the proposed solution, we compared using the ISIC-2019, ISIC-2019-2020, and ISIC datasets, which include both binary and multi-class classification. The proposed solution has also achieved better performance compared to several recent research works conducted to solve the challenges regarding skin cancer feature extraction and hence improve cancer detection accuracy. We have demonstrated that the proposed framework effectively focuses on clinically relevant lesion regions using Grad-CAM as an explainable AI (XAI) technique.

KEYWORDS: ResNet-50; self-attention; particle swarm optimization; explainable AI; skin cancer classification

1 Introduction

Early detection of skin cancer can increase survival rates among affected individuals. It is challenging to differentiate between malignant and benign cancer because their visual characteristics are similar [1]. This highlights the urgent need for computer-aided diagnosis (CAD) tools to support accurate and timely diagnoses [2]. CAD systems can aid the clinicians to enhance diagnostic accuracy and decrease the time required for diagnosis. The images of skin lesions are unique and complex, for example, color variations, skin texture, shape of the skin lesion, and the irregular border of the tumor, pose significant challenges

in distinguishing benign from malignant lesions [3]. The normal CAD workflow comprises of several steps: image preprocessing, feature extraction, segmentation, and classification [4]. Moreover, traditional CAD techniques face several challenges in preprocessing dermoscopic images acquired from individuals with diverse skin tones and ethnic backgrounds [5]. Furthermore, CAD systems rely on domain expertise for feature extraction and detection of malignant tumors. In addition, dermoscopic images often contain artifacts, such as uneven illumination, veins, hair, air bubbles, and color calibration marks, which reduce the accuracy of feature extraction and adversely affect the diagnostic process [6–8]. Therefore, automated medical image analysis using deep learning models has gained significant attention [9]. Although these models excel in identifying malignant and benign skin cancer, data imbalance remains a critical challenge for accurate skin cancer classification. Training on imbalanced datasets reduces the generalization ability of deep learning models, thereby limiting their classification performance.

Most deep learning models for medical image analysis, such as skin cancer classification, are built using dozens of Convolutional Neural Network (CNN) layers including, ResNet-50, ResNet-101, VGG-16, MobileNet, InceptionV3, Xception, EfficientNetB4, DenseNet, and others [9,10]. Data augmentation was used to overcome the generalizability issue (overfitting) due to an imbalanced training dataset. The augmented features were then processed through a stacked CNN architecture for skin cancer classification [11]. In the diagnosis of keratinocyte carcinoma, the authors [12] conducted comparative studies between AlexNet and several ResNet-50 variants. Among these models, ResNet-50 was identified as the most effective model [12]. Comparative studies have also evaluated different activation functions for binary classification, while a transfer learning approach using the ResNet-50 model has been investigated for multi-class skin cancer classification [13,14]. In addition, ensemble methods, including different optimization methods such as the sand cat swarm optimization, have also been experimented within the research [15–17]. Feature extraction also complicates the task because capturing the nuanced differences in skin texture, color, complexion, and shape inherent in dermoscopic images requires highly optimized architectures and robust annotated data. A hybrid approach combining transfer learning and zero-shot learning, Lightweight Vision Transformer (LViT), ensemble deep learning, and DEnSha on the Human Against Machine with 10,000 training images (HAM10000) dataset improved the recognition of rare lesions [18–21]. On the International Skin Imaging Collaboration (ISIC)-2019, CNN backbones such as InceptionV3, Xception, and DenseNet201 have improved classification through data preprocessing and augmentation [22]. Hybrid Convolutional Neural Network–Artificial Neural Network (CNN–ANN) and Convolutional Neural Network–Random Forest (CNN–RF) models combined with Principal Component Analysis (PCA)-based feature fusion have also improved melanoma detection [23]. Using transfer learning alone, a fine-tuned ResNet-50 model reached 93.5% accuracy in ISIC [24]. A hybrid InceptionV3–DenseNet121 model reached over 92% accuracy [25] and a weighted ensemble of CNNs achieved 93.36% accuracy for melanoma classification [15]. In recent years, the classification of skin cancer has also been further improved by using attention mechanisms, hybrid architectures, and multiple modal fusion. Transfer learning classifiers trained with attention mechanisms, based on Xception [26] and Scaled Dot Attention Mechanism (SDAM)-UNet [27] have shown better performance for localization of lesions in HAM10000 and ISIC datasets. These hybrid architectures that combine CNNs with Swin Transformers [28] and ensemble transfer learning models that incorporate patient metadata and dermoscopic images [29,30] have further enhanced classification accuracy. The generalizability of transfer learning for seven class skin lesion classification was confirmed in a large scale benchmarking study on different CNN architectures [31].

Despite recent advances in CNN architectures, the classification of skin cancer remains difficult owing to the visual similarity between skin lesions, the extreme imbalance of datasets, and the lack of discriminative features in dermoscopic images. Furthermore, artifacts such as hair, lighting variations, and skin textures

cause poor model generalization and make automated diagnosis less reliable. In this research, a framework is proposed to improve feature representation, stability, and generalization. The proposed solution is based on a ResNet-50 architecture that is pre-trained on the ImageNet dataset and enhanced with a self-attention mechanism to help focus on diagnostically relevant parts of the image, while suppressing irrelevant information from the background. A dual pooling strategy is used that includes max pooling and global average pooling to get local contextual features and global contextual features, which are necessary for discriminating similar patterns of lesions (as shown in Fig. 1). The normal kernel initializer, L2-regularization, dropout, early stopping, and ReduceLROnPlateau based adaptive learning rate scheduler are used as regularizations for stable convergence and to avoid overfitting. The Particle Swarm Optimization (PSO) algorithm automatically tunes some of the crucial parameters namely Learning rate, Dropout rate, and Regularization factor to ensure convergence and performance stability. Data augmentation and transfer learning are applied to address data imbalance, enhancing model robustness in various skin lesion classes. The proposed model was trained and tested on various publicly available datasets with 10-fold cross validation with binary classification as well as multiclass classification problems, such as HAM10000, ISIC-2019, ISIC-2019-2020, and ISIC. The results showed that the framework performed better than the other frameworks in terms of generalization, high accuracy in most of the datasets, and therefore, it is efficient in automated skin cancer classification.

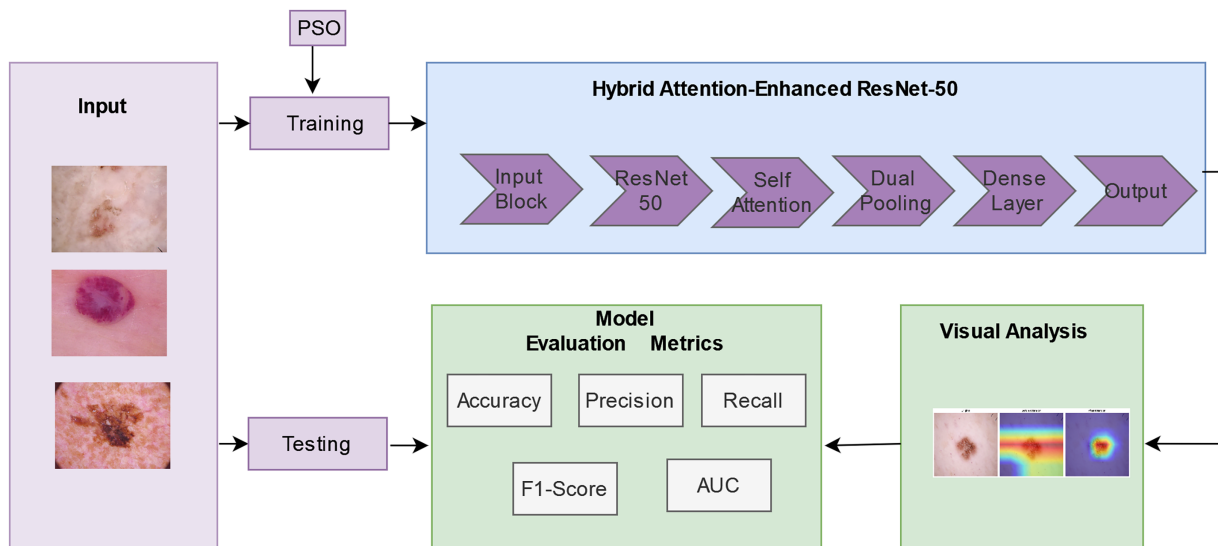


Figure 1: Block diagram of the proposed system.

We propose a framework to overcome the challenges of accurate and reliable classification of skin cancer in dermoscopic images, addressing issues of representation of features, training stability, and transferability of the model between datasets. This work makes the following key contributions:

1. **Integration of Transfer Learning with Self-Attention:** We have developed a framework that combines transfer learning and a self-attention mechanism on a ResNet-50 backbone pre-trained on ImageNet. Therefore, the model can pay more attention to the discriminative lesion regions and less to any background noise and irrelevant artifacts.
2. **Dual Pooling Strategy for Rich Feature Extraction:** In the model, we have adopted both max pooling and global average pooling to learn the most informative spatial and contextual features. Therefore, both fine-grained local texture and global lesion-level context are jointly preserved for robust multi-class skin lesion classification.

3. **Comprehensive Training Optimization:** L2-regularization, dropout, He normal initialization, early stopping, and ReduceLROnPlateau were used to stabilize the convergence and thus avoid overfitting.
4. **Particle Swarm Optimization (PSO)-based hyperparameter tuning:** The PSO technique optimizes critical parameters like learning rate, dropout rate and, regularization factor for better convergence and stability in performance.
5. **Cross-Dataset Validation for Robustness:** We have validated with several datasets such as HAM10000, ISIC-2019, ISIC-2019-2020, and ISIC. The cross-dataset evaluation, both for binary and multi-class tasks, offers a full analysis of the generalization ability and clinical usefulness of the framework.
6. **Superior Performance over State-of-the-Art Models:** The proposed framework has achieved better performance compared to several recent research studies conducted to solve the challenges regarding skin cancer feature extraction and hence improving the cancer detection accuracy.
7. **Explainability and Visual Interpretation:** We have used explainable AI technique like Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the model's attention and feature importance.

The remainder of this paper is organized as follows. [Section 2](#) reviews the related work on skin-lesion analysis. [Section 3](#) presents the proposed methodology, including the Hybrid Attention-Enhanced ResNet-50 architecture, hyperparameter optimization using Particle Swarm Optimization (PSO), and Experiment Setup. [Section 4](#) reports the experimental results and provides a detailed discussion, including an ablation study, computational complexity analysis, visual interpretation, and explainability. Finally, [Section 5](#) concludes the study.

2 Related Work

Skin cancer is a fatal disease, and currently studies are being conducted to classify this disease [9]. Human skin texture, including the complexion and overlapping features of this disease presented in images, are very challenging for discriminating different categories of skin cancers by trained deep learning models. Newly researched deep learning models are trying to overcome these challenges.

The authors [10] experimented with the ResNet-50 deep learning model to classify different skin cancers by incorporating complex categories of skin lesions. Imbalance skin cancer classification image dataset causes overfitting of trained CNN models. The authors [11] balanced the dataset using data augmentation and trained stacked CNN models on it, which achieved 0.73 accuracy score during validation. The authors developed an automated keratinocyte carcinoma skin cancer detection method by comparing Alex-Net, and different ResNet-50 architectures [12]. Activation functions at the output layers produce confidence scores based on which the classification output is determined. Hence, the authors compared the impact of applying Rectified Linear Unit (ReLU), softmax, and sigmoid activation in CNN model and Dense CNN model. Their experiment showed that the Sigmoid activation function performed well by scoring 0.76 and 0.80 accuracy, respectively [13].

Dermatologists cannot detect overlapping skin cancer features well when diagnosing different categories of skin cancer. For multiclass skin cancer classification, the authors proposed a transfer learning based ResNet-50 classification method [14]. The features that significantly contribute to the identification of different types of skin cancers are extracted by fine tuning the pre-trained ResNet-50 model and given as input to the XGBoost skin cancer classification model [17]. Segmentation helps extract important lesions in dermoscopic images by removing irrelevant backgrounds. The authors extracted and reduced features using segmentation and input these key features into the ensemble model constructed with different CNN models [32]. Early detection of skin cancer is very complicated and costly. The authors addressed this problem by leveraging deep learning methods and proposed the sand cat swarm optimization with

ResNet-50 (SCSO-ResNet-50) method to enhance the latent feature identification as well as dimensionality reduction [16].

CAD based skin cancer diagnosis has different phases including, dermoscopic image collection, gaussian filtering, segmentation, feature extraction, and finally skin cancer classification. The authors proposed an ensemble approach that includes different types of deep learning model for binary classification of malignant melanoma and non-cancerous skin ailments [15]. However, increasing the number of classes in a dataset can badly affect the performance of a skin cancer classification model. The authors proposed a multi-classification ensemble model of different CNN architectures and different segmentation techniques using U-Net and RP-Net models, which are integrated by Jaccard similarity-based fusion model [32]. The authors introduced a method to extract features by using ResNet-50 pre-trained model, then the features are compressed by the Block Truncation Code technique. The compressed features are then given as input to the downstream classifiers for better classification [33].

The authors tested different variations of self-attention. The first is pairwise self-attention, which improves the traditional attention mechanism, and the next is patch-wise self-attention, which is superior to convolution layers. Suggested that patch-wise self-attention-based models show better performance with different datasets in computer vision applications [34]. Magnetic Resonance Imaging (MRI) is a powerful imaging technique, but has a limitation of slow acquisition capability. The authors developed SAT-Net, a deep learning-based model that adds self-attention to the convolution, algorithm that reduces time without damaging information [35]. Oral and dental disease can be a cause of cancer that can deteriorate overall health. The authors have proposed a novel self-attention technique that integrates features at the feature level by combining efficientNetB0 and efficientNetB1, which improved the performance compared to previous research based on CNN [36]. The authors experimented on self-attention within ResNet for skin cancer classification, but the performance of the model is low [37].

The authors combining transfer learning with zero-shot learning on the HAM10000 dataset improved recognition of rare and unseen skin lesions through semantic embedding-based feature extraction. This hybrid approach enhanced accuracy and generalization, which could be useful for improving automated dermatological diagnosis [18]. The authors showed that data preprocessing and augmentation using pre-trained CNNs such as InceptionV3, Xception, and DenseNet201 effectively improved classification accuracy, highlighting the importance of balanced and well-prepared datasets in skin cancer detection [22]. Hybrid CNN-based models that combine AlexNet, GoogLeNet, and VGG16 with ANN and Random Forest were applied to the ISIC-2019 dataset, where PCA-based feature fusion and Geodesic Active Contour (GAC) segmentation enhanced melanoma detection accuracy [23]. Five pre-trained transfer learning models were applied to the ISIC dataset for binary skin cancer detection, where fine-tuning and data augmentation enhanced performance. Among them, ResNet-50 achieved the highest accuracy of 93.5%, demonstrating its effectiveness for automated diagnosis [24]. A hybrid deep learning model combining InceptionV3 and DenseNet121, with dataset preprocessing and weighted fusion, improved performance, achieving over 92% accuracy in skin cancer diagnosis [25]. Additionally, a weighted ensemble of deep learning models, including InceptionV3, VGG16, Xception, and ResNet-50, achieved 93.36% accuracy on the ISIC dataset, demonstrating the effectiveness of ensemble learning for melanoma detection [15]. A study proposed an Efficient Dual Attention-based ResNet-50 (EDA-ResNet-50) integrating multi-scale features with channel and spatial attention, achieving 93.18% accuracy for skin cancer classification with improved efficiency and interpretability [38]. In this study, the authors have proposed LViT, a transfer learning and lightweight B-16 vision models have improved accuracy, helping overcome visual similarity among cancer types and improving early detection reliability in large and diverse dermatological image datasets [19]. The authors proposed a hybrid model that combines preprocessing techniques with deep-learning architectures to improve

skin-lesion segmentation. Hair-removal methods and ensemble networks gain accuracy on HAM10K and ISIC datasets, indicating the effectiveness of hybrid approaches for precise melanoma boundary detection and analysis [20]. The author presented a four-approach based discriminant feature selection model called Deep, Ensemble, and Shallow (DEnSha) Neural Networks approach. These features are bundled together with ensemble techniques and trained with multi-class Support Vector Machine (SVM) and shallow neural networks, and both were found to be very accurate and robust on ISIC 2018 data [21].

Recent advances in transfer learning have improved skin cancer detection accuracy. The authors introduce an attention-enhanced CNN model emphasizing the Xception algorithm, where attention mechanisms inside the CNN preferentially concentrate on critical lesion regions to strengthen feature extraction and improve classification accuracy, assessed on the HAM10000 dataset across seven categories of skin lesions [26]. Building on attention-based approaches, an advanced AI-driven framework integrating a Scaled Dot Attention Mechanism (SDAM) with a modified UNet architecture was proposed for skin lesion detection, evaluated on the HAM10000 dataset and further validated on the ISIC and PH2 datasets. The SDAM, applied between the encoder and decoder stages, enabled the model to prioritize relevant lesion regions while suppressing background noise, achieving strong generalization across multiple skin lesion types [27]. The integration of patient metadata with dermoscopic images has emerged as an important direction for improving skin cancer diagnosis. A transfer learning-based ensemble framework leveraging ResNet50, Xception, MobileNet, EfficientNetB0, and DenseNet121 on the ISIC dataset showed that fusing patient metadata with skin lesion images significantly improved classification accuracy compared to using images alone [29]. This finding was further confirmed in 2026, where deep learning fusion of structured patient metadata with dermoscopic images substantially boosted classification performance on a large clinical dataset [30]. A novel hybrid deep learning model named SkinEHDLF was proposed integrating ConvNeXt, EfficientNetV2, and Swin Transformer architectures with an adaptive attention-based feature fusion mechanism, fine-tuned using transfer learning on the ISIC 2024 dataset. This hybrid methodology combined the proficient feature extraction of CNNs with the long-range attention of Swin Transformers to achieve highly accurate and reliable lesion classification [28]. A comprehensive benchmarking study evaluated 38 CNN architectures from ten families including ResNet, EfficientNet, DenseNet, MobileNet, and Xception using transfer learning on the HAM10000 dataset for seven-class skin lesion classification. The models were further evaluated on the ISIC-2019 dataset confirming their robustness [31]. In resource-limited healthcare settings, VGG16 and ResNet-50 were compared for binary classification using locally collected clinical images from India, demonstrating the practical applicability of transfer learning models in resource-constrained environments [39]. Furthermore, a multi-modal skin lesion classification framework for early cancer diagnosis demonstrated that multi-modal fusion of deep image features with clinical data on the HAM10000 dataset significantly improves early melanoma diagnosis [40].

The reviewed articles addressed the challenges of skin cancer classification tasks, which include overlapping features, imbalanced datasets, feature extraction, and low generalizability. In our research, we proposed a model for identifying overlapping features and increasing the accuracy of multiclass classification. Our proposed solution focused on balancing the image dataset by data augmentation, transfer learning from pre-trained model by fine tuning with new smaller dataset, and more precise feature extraction by determining feature context using self-attention technique. This approach will yield better accuracy because the model would be able to focus on relevant skin cancer feature regions. The proposed solution could achieve higher accuracy for multiclass skin cancer detection while reducing false positives in overlapping feature regions.

3 Methodology

3.1 Dataset

In this research, we evaluated our solution using both multi-classification and binary classification datasets. The Human Against Machine with 10000 training images (HAM10000) dataset is a collection of dermoscopic images of different pigmented skin cancer lesions [41]. The dataset is available on the website of The Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia. It comprises 10,015 high resolution images collected over 20 years. The dataset is divided into 7 categories. The different categories of skin lesion classes are Actinic keratoses (AKIEC), Basal cell carcinoma (BCC), Benign keratosis-like lesions (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic nevus (NV), and Vascular lesions (VASC). Fig. 2 shows samples from these datasets.

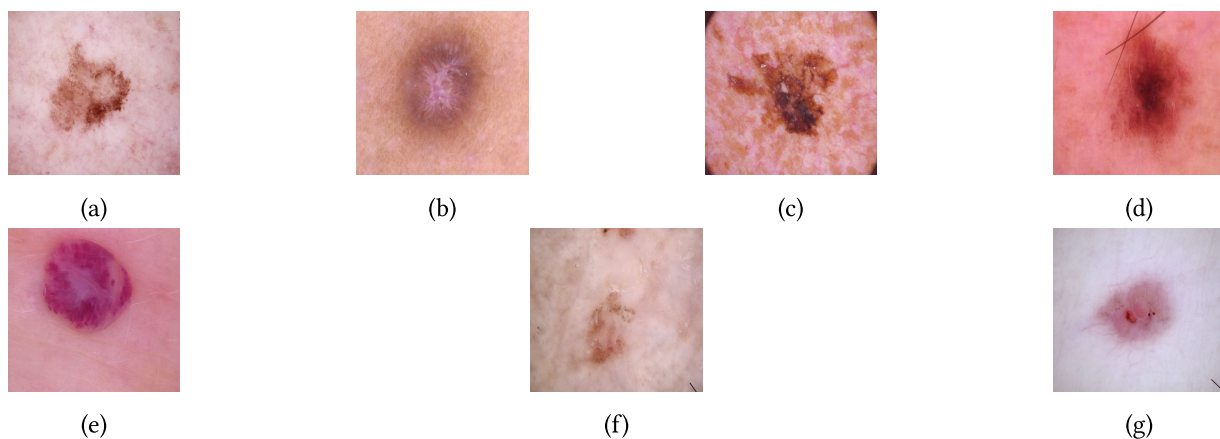


Figure 2: Sample HAM10000 dataset of different classes (a) Benign keratosis-like lesions (BKL), (b) Dermatofibroma (DF), (c) Melanoma (MEL), (d) Melanocytic nevus (NV), (e) Vascular lesions (VASC), (f) Actinic keratoses (AKIEC), and (g) Basal cell carcinoma (BCC).

The next dataset is the (ISIC-2019) challenge dataset of dermoscopic images, developed by the International Skin Imaging Collaboration (ISIC) [42]. This dataset comprises the previous years both ISIC 2018 and 2017. 25,331 skin cancer images are available. It is labeled into 8 classes like Melanocytic nevus (NV), Basal cell carcinoma (BCC), Dermatofibroma (DF), Melanoma (MEL), Actinic keratosis (AK), Benign keratosis-like lesions (BKL), Vascular lesion (VASC), and Squamous cell carcinoma (SCC). Fig. 3 shows samples from these datasets.

The next two datasets are binary classification datasets. The first ISIC-2019-2020 dataset, which combines both ISIC-2019 and ISIC-2020 challenge dataset [43]. The dataset comprises 11,449 images of skin lesions, grouped into two classes. Of which 5106 images belong to mel and 6343 images is of nevus. Fig. 4 are the sample images showcasing variations in texture, colour, and lesion patterns.

Another is ISIC binary classification dataset, consists of 3297 images of skin diseases, which has 1800 malignant and 1497 benign classes images. This dataset title “Malignant vs. Benign Skin Cancer” was downloaded from Kaggle. The benchmark dataset uses images categorized into benign and malignant melanocytic lesions that come from the ISIC archive [44]. Fig. 5 are the sample from these datasets.

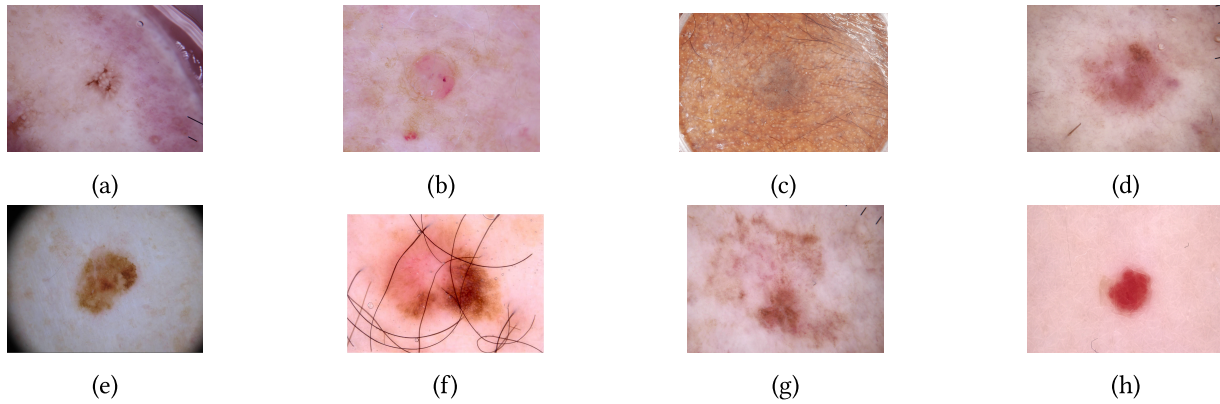


Figure 3: The sample from ISIC-2019 dataset with multiple classes (a) Actinic keratosis (AK), (b) Basal cell carcinoma (BCC), (c) Benign keratosis-like lesions (BKL), (d) Dermatofibroma (DF), (e) Melanoma (MEL), (f) Melanocytic nevus (NV), (g) Squamous cell carcinoma (SCC), (h) Vascular lesion (VASC).

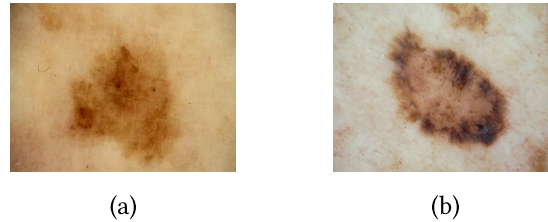


Figure 4: (a) Nevus class and (b) Mel class from ISIC-2019-2020 dataset.

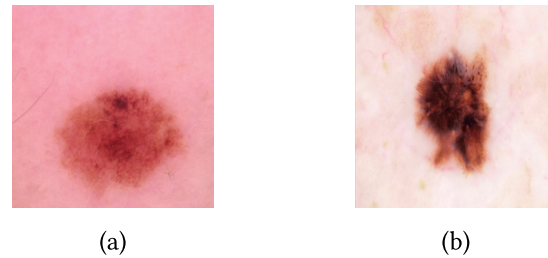


Figure 5: (a) Benign class and (b) Malignant class from ISIC dataset.

Fig. 6 presents the class distribution across the HAM10000, ISIC-2019, ISIC 2019–2020, and ISIC datasets. It is observed that the HAM10000 and ISIC-2019 datasets exhibit a pronounced class imbalance, predominantly in the NV category, whereas the binary datasets maintain a relatively balanced distribution between classes. Details of datasets are shown in Table 1.

3.2 Data Preprocessing

A number of augmentation techniques, like rotation, flipping, scaling, translation, shearing, zooming, brightening, contrast adjustment, and noise addition are used after the images are resized to 224×224 . In order to get a reliable performance evaluation, we used 10-fold cross validation. The dataset was then split into ten folds: nine for training and one for validation, and this process was repeated ten times. Across 10 independent runs, we computed the mean and standard deviation for each evaluation metric.

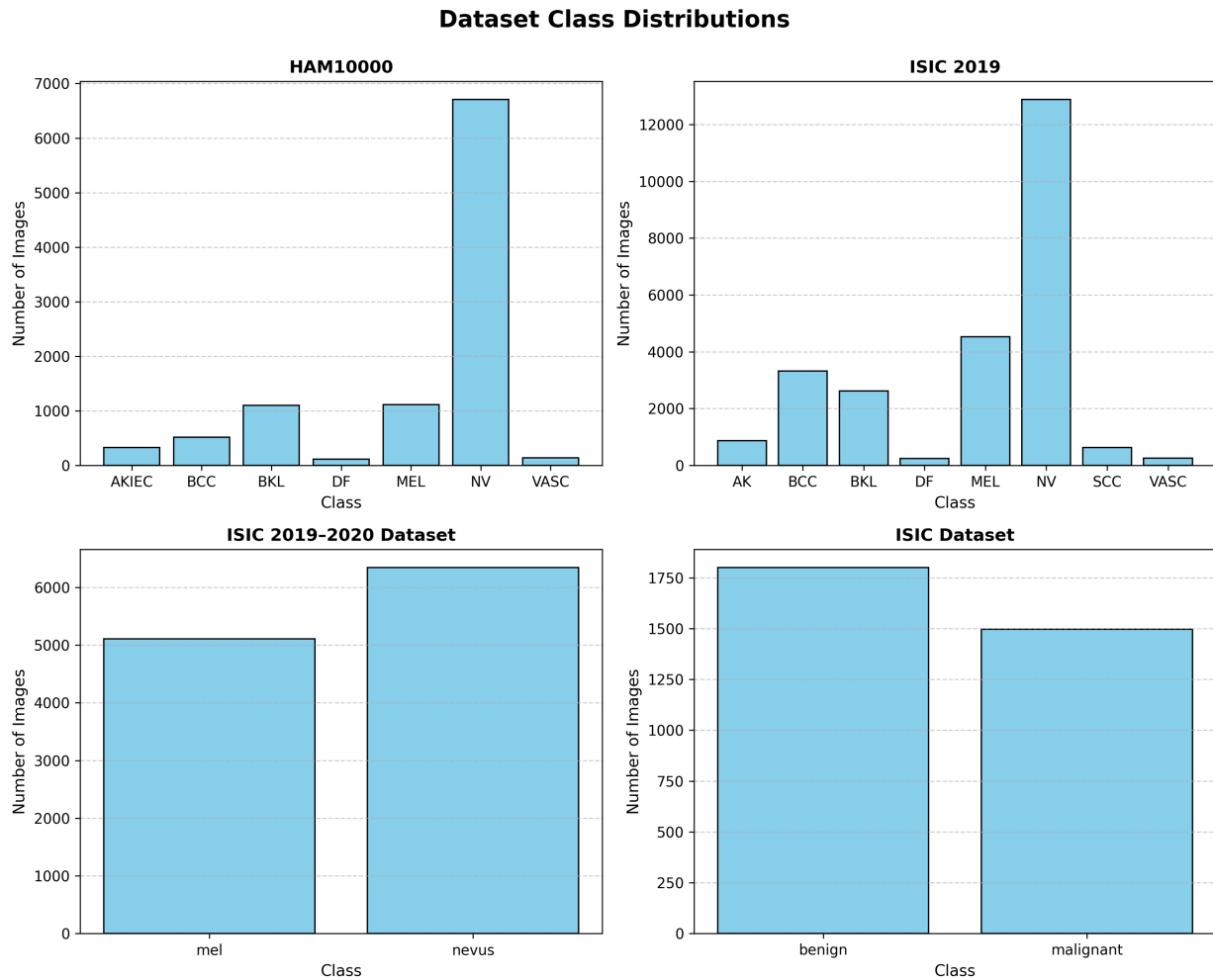


Figure 6: Class distribution for HAM10000, ISIC-2019, ISIC 2019–2020, and ISIC datasets.

Table 1: Dataset details.

Dataset	Total Images	Class
HAM10000	10,015	7
ISIC-2019	25,331	8
ISIC-2019-2020	11,449	2
ISIC	3297	2

3.3 Hybrid Attention-Enhanced ResNet-50 model Architecture

The pretrained ResNet-50 model is enhanced with self-attention for feature extraction in computer vision, which is suitable for identifying long-range patterns and generalizing complex feature representations [45]. Here, we have proposed a model for discriminating overlapping features and retaining low-level features, including efficient generalization of unseen data, as shown in Fig. 7. To preserve feature information, it uses residual connections and enables efficient learning to prevent issues like vanishing gradient. The residual block is the main block with skip connections that allow the input layer to bypass one or more layers and add to the output layer. There are stacks of residual blocks together in ResNet-50. Instead of training the

layers directly to learn the desired output $H(x)$, the network is designed to learn residual mapping $F(x)$. This residual mapping represents the difference between the desired output $H(x)$ and input x .

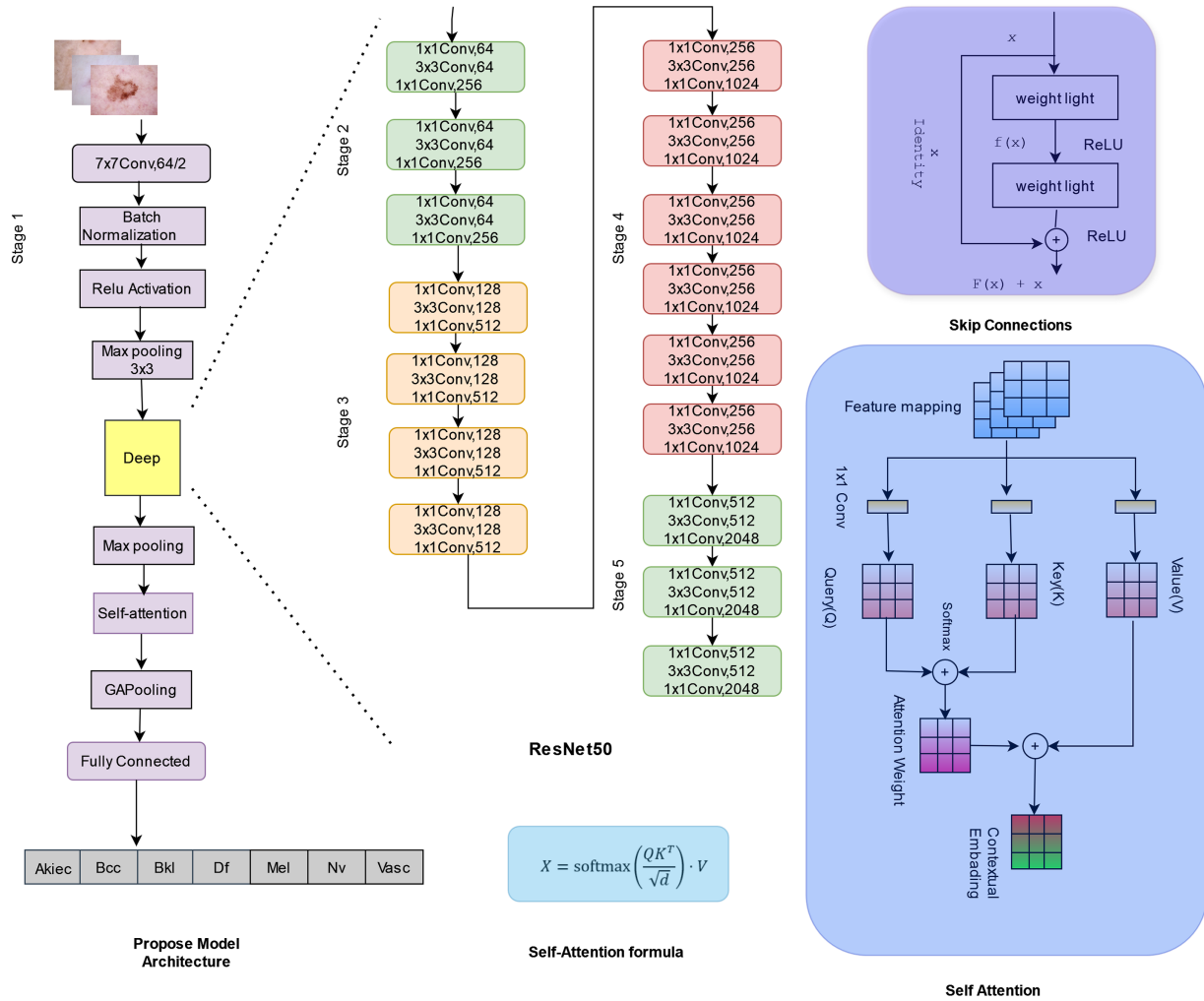


Figure 7: Architecture of the proposed model.

The mapping can be represented as

$$H(x) = F(x) - x \quad (1)$$

This gives the residual mapping is

$$F(x) = H(x) - x \quad (2)$$

One of the primary advantages of a residual network is that it allows the network to bypass layers that might negatively affect the overall performance. By minimizing the dominance of the ineffective layer, the residual connections ensure that the network focuses on learning meaningful features. The hierarchical structure allows extraction at different levels. The lower level extracts patterns like edges, texture, and lines. In the intermediate level concentrate on shape and contours and more complex spatial arrangements. The deep level learns contextual representations such as concept of object and its relationships with each other.

The max-pooling layer is connected between the ResNet-50 layer and the self-attention layer, which will diminish the spatial size of the feature map. Therefore, applying self-attention after max-pooling helps the model to focus on important relations between spatial regions.

If the input notation is represented as

$$X \in \mathbb{R}^{H \times W \times C} \quad (3)$$

where W , H are spatial dimensions and C is the number of channels.

Max-Pooling reduces the spatial dimension to $H' \times W'$ by retaining dominant values in local regions.

$$X_{\text{pool}} \in \mathbb{R}^{H' \times W' \times C} \quad (4)$$

Self-attention captures global relationships between spatial regions to produce a contextualized feature map.

$$X_{\text{attn}} \in \mathbb{R}^{H' \times W' \times C} \quad (5)$$

The placement of the self-attention module after the max-pooling layer is a deliberate architectural decision. By first applying max pooling to compress the feature maps from ResNet-50 Stage 4, the spatial resolution is reduced, allowing the self-attention mechanism to compute global relationships between spatially distant lesion regions more efficiently and with lower computational overhead. This design ensures that the attention module focuses on semantically rich, dominant local features rather than raw high-resolution spatial noise, thereby improving both discriminative power and convergence stability. Following the self-attention module, global average pooling is applied to consolidate the attention-weighted feature maps into a compact global representation, capturing the overall contextual relationships across the lesion region before passing the features to the fully connected classification layer.

Then the maximum value in each pooling window

$$X_{\text{pooling}}(i, j, c) = \max_{(m,n) \in \text{window}(i,j)} X(m, n, c) \quad (6)$$

where (i, j) indexes the pooled spatial grid and C indexes the channel. The window (i, j) defines the spatial region covered by the pooling kernel. X_{pool} represents prominent local features in a compact spatial format.

Self-Attention operates on X_{pool}

$$X_{\text{pooling}} \in \mathbb{R}^{N \times C}, \quad N = H' \times W' \quad (7)$$

In attention mechanism linear projection represented as

$$Q = X_{\text{pool}} W'_Q, \quad K = X_{\text{pool}} W'_K, \quad V = X_{\text{pool}} W'_V \quad (8)$$

where $W'_Q, W'_K, W'_V \in \mathbb{R}^{C \times d}$ are the learnable weights.

The attention weights compute attention scores and normalize.

$$A = \text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right), \quad \text{where } A \in \mathbb{R}^{N \times N} \quad (9)$$

defines the relationship between all spatial regions.

The contextualized feature maps aggregate features using attention weights.

$$X_{\text{attn}} = AV \quad (10)$$

$X_{\text{attn}} \in \mathbb{R}^{N \times d}$ represents globally contextualized features.

$$X_{\text{attn}} = \text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right) \cdot V \quad (11)$$

The SoftMax ensures $A \in \mathbb{R}^{N \times N}$ is a probabilistic matrix where rows sum to 1.

Replacing Q , K , and V

$$X_{\text{attn}} = \text{softmax} \left(\frac{(X_{\text{pool}} W_Q)(X_{\text{pool}} W_K)^T}{\sqrt{d}} \right) \cdot (X_{\text{pool}} W_V) \quad (12)$$

The output X_{attn} is a globally contextualized feature map that combines local prominence (from pooling) and global context (from attention).

Next, to compute the average of all spatial values on each channel, a global average pooling (GAP) is applied to the self-attention feature map X_{attn} .

$$\text{GAP}(X) = \frac{1}{N} \sum_{i=1}^N X_{\text{attn}}[i, :] \quad (13)$$

where $N = H' \times W'$ are the flattened spatial dimensions. $X_{\text{attn}}[i, :]$ denotes the i -th spatial location across all d channels.

$$\text{GAP}(X) = \frac{1}{N} \sum_{i=1}^N (A \cdot V)[i, :] \quad (14)$$

$$\text{GAP}(X) = \frac{1}{N} \sum_{i=1}^N \left[\text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right) X_{\text{pool}} W_V \right][i, :] \quad (15)$$

Factorize attention weights are row-normalized, and summing over all rows shows the global effect of V .

$$\text{GAP}(X) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N A[i, j] V[j, :] \quad (16)$$

$$\text{GAP}(X) = \frac{1}{N} \sum_{j=1}^N \left(V[j, :] \sum_{i=1}^N A[i, j] \right) \quad (17)$$

If the attention weights $\sum_{i=1}^N A[i, j]$ are uniform or normalized, the equation simplifies to

$$\text{GAP}(X) = \frac{1}{N} \sum_{j=1}^N V[j, :] \quad (18)$$

This means that GAP directly averages the feature values across all spatial regions without considering specific contributions. The final global average pooling results

$$\text{GAP}(X) = \frac{1}{N} \sum_{j=1}^N ((X_{\text{pool}} W_V)[j, :]) \quad (19)$$

To ensure the global features are normalized, apply an L2-normalization to the GAP output:

$$\text{GAP}_{\text{norm}}(X) = \frac{\text{GAP}(X)}{\|\text{GAP}(X)\|} \quad (20)$$

This transformation enhances the ability of the network to learn meaningful global features. When weights $\sum_{i=1}^N A[i, j]$ are included, GAP can focus on specific regions of the feature map, enhancing its contextual understanding. By applying W_V , the model adapts GAP outputs to capture more nuanced global features, improving its predictive performance.

3.4 Hyperparameter Optimization using Particle Swarm Optimization (PSO)

The selection of optimal hyperparameters significantly improves the stability and generalization of deep learning models. In this work, we adopt Particle Swarm Optimization (PSO) to automatically tune critical hyperparameters of the proposed model, such as learning rate (lr), dropout rate, and regularization factor. The PSO-based optimization ensures that the model attains high classification performance with minimal manual intervention.

Initially, a population (swarm) of N particles is randomly initialized within predefined bounds of the hyperparameter search space. Each particle represents a potential hyperparameter configuration $x_i = [lr, dropout, \lambda, \dots]$ and maintains a corresponding velocity v_i . The fitness of each particle is assessed by training the model with its respective hyperparameter set and computing the performance metric on a validation subset—primarily area under the curve (AUC), with secondary metrics including F1-score and accuracy.

Each particle retains its personal best position ($p_{best,i}$), while the swarm collectively tracks the global best (g_{best}). The velocity and position of each particle are updated according to:

$$v_i \leftarrow w \cdot v_i + c_1 r_1 (p_{best,i} - x_i) + c_2 r_2 (g_{best} - x_i) \quad (21)$$

$$x_i \leftarrow x_i + v_i \quad (22)$$

Here, w denotes the inertia weight controlling the balance between exploration and exploitation, and c_1 and c_2 are cognitive and social coefficients, respectively. The random variables $r_1, r_2 \sim U(0, 1)$ introduce stochasticity to encourage exploration of the search space. Particles are constrained within the defined search boundaries, and velocities are clamped to prevent divergence.

The optimization process continues until one of the termination conditions is met—either a predefined maximum number of iterations or convergence (no improvement in the validation metric for T consecutive iterations). The final best solution $x^* = g_{best}$ represents the optimal hyperparameter configuration, which is then used for retraining the model on the combined training and validation data, followed by evaluation on the test set.

PSO provides an efficient, derivative free approach for optimizing complex nonconvex objectives, making it well-suited for deep learning tasks where gradient based hyperparameter tuning is computationally expensive or unstable. By automatically discovering effective hyperparameter settings, the PSO-based optimization, as shown in Fig. 8, contributes to the improved accuracy, convergence speed, and generalization performance of the proposed model.

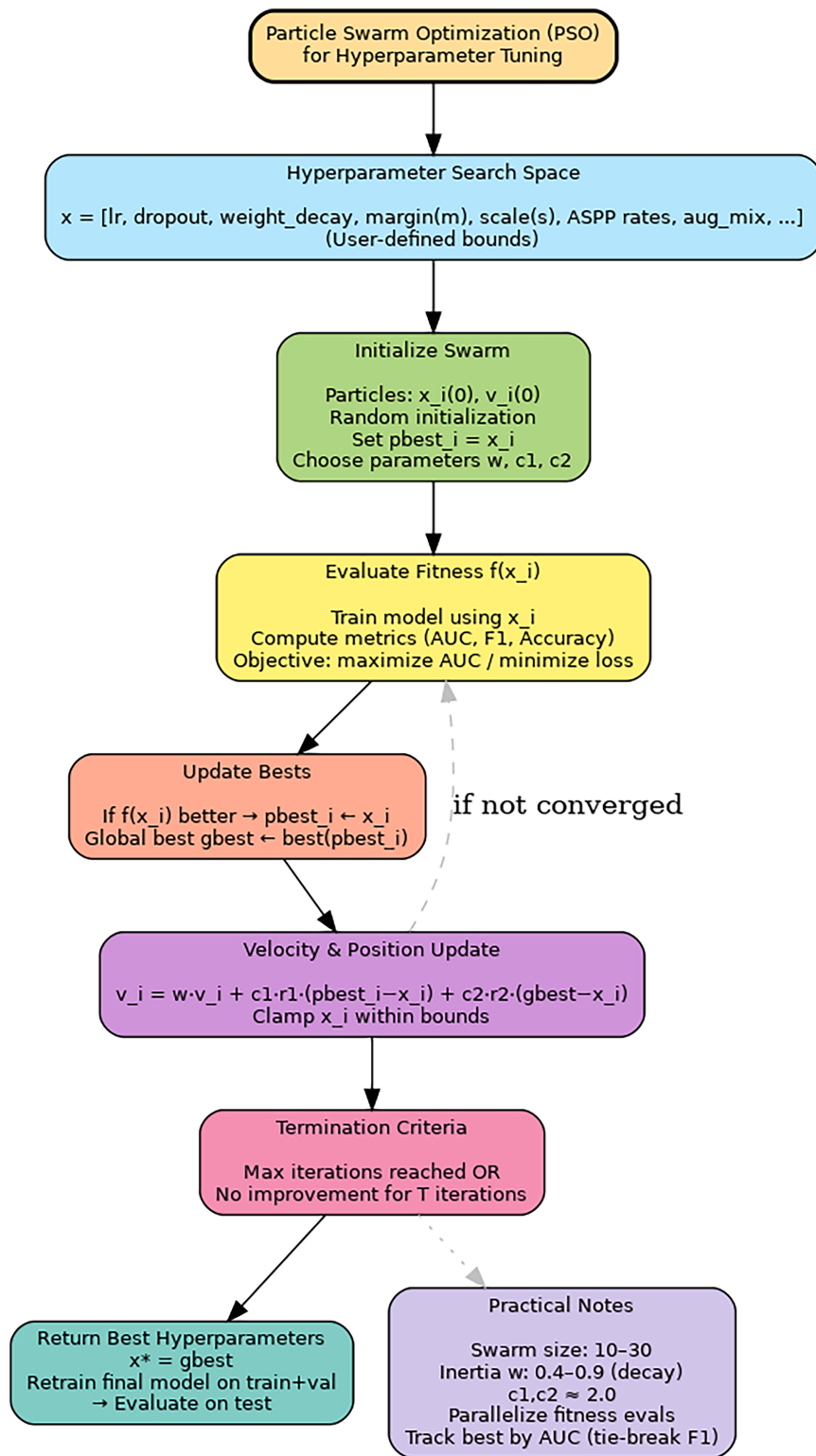


Figure 8: Flow diagram illustrating the Particle Swarm Optimization (PSO) process used for hyperparameter tuning in the proposed deep learning framework. Each particle explores the search space, updating its velocity and position based on both personal and global experience until convergence.

3.5 Algorithmic Integration of Particle Swarm Optimization with Attention-Enhanced ResNet-50

The Algorithm 1 illustrates how Particle Swarm Optimization (PSO) is integrated with the proposed Attention-Enhanced ResNet50 framework to achieve optimal hyperparameter tuning. Each particle in the swarm represents a potential configuration of hyperparameters such as learning rate, dropout rate, and regularization weight. The ResNet50 backbone, pretrained on ImageNet, is augmented with a self-attention module and a feature fusion mechanism combining max and average pooling to capture both local and global lesion characteristics. For every iteration, the model corresponding to each particle is trained and evaluated using the validation dataset to compute its fitness based on AUC and F1-score metrics. Particles then update their positions and velocities according to PSO dynamics, gradually converging toward the global best configuration that maximizes classification performance. Finally, the best-performing hyperparameter set is used to retrain the Attention-Enhanced ResNet50 model on the full training data, ensuring robust convergence and optimal generalization.

Algorithm 1: Integration of particle swarm optimization (PSO) in attention-enhanced ResNet50 framework

Require: Training dataset $\mathcal{D}_{train}$, Validation dataset $\mathcal{D}_{val}$, maximum iterations T_{max} , swarm size N , PSO parameters (w, c_1, c_2)

Ensure: Optimized hyperparameter configuration $\mathbf{x}^*$ for ResNet50

- 1: Initialize swarm of N particles: each particle i has a position (hyperparameter vector) $\mathbf{x}_i = [lr_i, dropout_i, \lambda_i]$ and velocity $\mathbf{v}_i$.
 - 2: Initialize personal best $pbest_i \leftarrow \mathbf{x}_i$ and global best $gbest \leftarrow \arg \max_i f(\mathbf{x}_i)$.
 - 3: **for** iteration $t = 1$ to T_{max} **do**
 - 4: **for** each particle $i = 1$ to N **do**
 - 5: Decode $\mathbf{x}_i$ to ResNet50 configuration:
Learning rate lr_i , Dropout rate $dropout_i$, Regularization λ_i .
 - 6: Construct Attention-Enhanced ResNet50:
Load pre-trained ResNet50 backbone.
Integrate Self-Attention block after Stage 4 feature maps.
Apply Max and Average pooling fusion.
Add Fully Connected (FC) layer with Softmax.
 - 7: Train the model with parameters $(lr_i, dropout_i, \lambda_i)$ using $\mathcal{D}_{train}$.
 - 8: Evaluate fitness $f(\mathbf{x}_i)$ on $\mathcal{D}_{val}$ using AUC (primary) and F1-score (secondary).
 - 9: **if** $f(\mathbf{x}_i) > f(pbest_i)$ **then**
 - 10: $pbest_i \leftarrow \mathbf{x}_i$
 - 11: **end if**
 - 12: **end for**
 - 13: Update global best: $gbest \leftarrow \arg \max_i f(pbest_i)$.
 - 14: **for** each particle $i = 1$ to N **do**
 - 15: Update velocity:
 $\mathbf{v}_i \leftarrow w \cdot \mathbf{v}_i + c_1 r_1 (pbest_i - \mathbf{v}_i) + c_2 r_2 (gbest - \mathbf{x}_i)$
 - 16: Update position:
 $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$
 - 17: Clamp $\mathbf{x}_i$ to valid search space bounds.
 - 18: **end for**
 - 19: **end for**
 - 20: **return** Best configuration $\mathbf{x}^* = gbest$
 - 21: Retrain final Attention-Enhanced ResNet-50 using $\mathbf{x}^*$ on full $\mathcal{D}_{train} \cup \mathcal{D}_{val}$.
 - 22: Evaluate final model on test dataset $\mathcal{D}_{test}$.
-

3.6 Experiment Setup

In this experiment, a graphics processing unit (GPU) based system, Tesla V100-SXM2-32GB was used to provide high computational performance. Categorical cross-entropy loss combined with the Adam optimizer and a dynamic learning rate schedule was employed. Early stopping was applied to halt training when no improvement was observed. The model adopted SoftMax for multi-class recognition tasks while employing Sigmoid for binary recognition functions. This configuration demonstrates the ability to manage both binary and multi-class classification tasks with steady convergence and optimal model performance. To address the challenges of skin cancer classification, the performance of ten different deep learning models is compared. The first comparison is done with different pretrained models like VGG16, MobileNet, InceptionV3, Xception, EfficientNetB4, DenseNet, InceptionResNetv2, and ResNet-50. This comparison is conducted using HAM10000 dataset that contains seven classes. In the second analysis, the robustness of the proposed solution is investigated using three different ISIC dataset is used. [Table 2](#) shows the PSO optimized hyperparameter used for the proposed solution.

Table 2: PSO optimized hyperparameter used for the proposed solution.

Hyperparameter	Values
Adaptive LR	ReduceLROnPlateau
Batch	64
Dropout	0.4
Loss	Categorical Cross-Entropy
Regularization	10^{-2}
Epochs	25–60
K-Fold	10
Early Stop	val_loss

4 Results and Discussion

4.1 Results

In this study, we examined the performance of various models based on Precision, Recall, Accuracy, and F1-score.

To ensure a reliable estimate of model performance, a 10-fold cross-validation technique was utilized throughout the dataset. [Table 3](#) shows the performance metrics for all the models trained on the HAM10000 dataset. Starting with the Base-CNN architecture with four CNN layers, each followed by max-pooling, achieves an accuracy of 53% and f1-score of 32%, which indicates model is unable to generalize the classes. Next, VGG16, a 16 layers deep convolutional neural network, achieved a maximum accuracy of 65%, an F1-score of 47%, a precision of 43%, and a recall of 57%, showing a slight improvement. The lightweight MobileNet, which utilizes depth-wise separable convolution, shows an accuracy of 85%, an F1-score of 78%, a precision of 75%, and a recall of 82%. InceptionV3, capable of capturing the complex patterns of skin cancer, achieved precision, recall, accuracy, and F1-scores ranging from 76% to 86%. Similarly, the Xception algorithm achieved the highest performance metrics (precision, recall, accuracy, and F1-scores) ranging from 77% to 87%. EfficientNetB4, a convolutional neural network that employs a scaling technique to balance width, depth, and resolution for increased efficiency, achieved a maximum accuracy of 87%, an F1-score of 79%, a precision of 76%, and a recall of 84%. DenseNet contains multiple dense blocks, and in each block convolution layer is connected to all previous layer has maximum accuracy of 89% and F1-score of 77%. The InceptionResNet architecture is a combination of Inception and ResNet model achieved maximum accuracy of 89%, F1-score of 80%, precision of 79%, and recall of 82%. ResNet-50 consists of 50 layers of convolution

neural network, which uses residual learning with skip connection to avoid the problem of vanishing gradient achieved maximum accuracy obtain of 90%, f1-score of 90%, precision of 91%, and recall of 89%. Finally, the proposed solution outperforms all the models with accuracy of 97%, precision of 98%, recall of 97%, and f1-score of 97%.

Table 3: Comparing performance of baseline models with proposed solution trained on HAM10000 dataset.

Model	Precision (%)	Recall (%)	Accuracy (%)	F1-Score (%)
Base-CNN	28	38	53	32
VGG16 [46]	43	57	65	47
MobileNet [47]	75	82	85	78
InceptionV3	76	81	86	78
Xception	77	79	87	77
EfficientNetB4 [48]	76	84	87	79
Stack CNN [11]	77	–	77	74
DenseNet [49]	81	75	89	77
InceptionResNet	79	82	89	80
ResNet50 [10]	91	89	90	90
Proposed Solution	98	97	97	97

Fig. 9 showcases the deep learning model presented in Table 3 comparison study on HAM10000 dataset. The x -axis displays the different models, while y -axis is the percentage. The blue, yellow, green, and red colours represent precision, recall, accuracy, and F1-score, respectively. The Base-CNN has the lowest performance rates, as it is unable to grab the complex pattern effectively. A minor improvement seen in VGG16. MobileNet, InceptionV3, Xception, EfficientNetB4, and InceptionResNet exhibit almost equal executive capabilities, making these models fit for classification works. As it can be seen, ResNet-50 is the best-performing algorithm with respect to precision, recall, accuracy, and F1-score. The proposed solution successfully outperforms every architecture by achieving superior scores in all operational metrics, thus establishing its ability to extract features effectively and perform precise classifications.

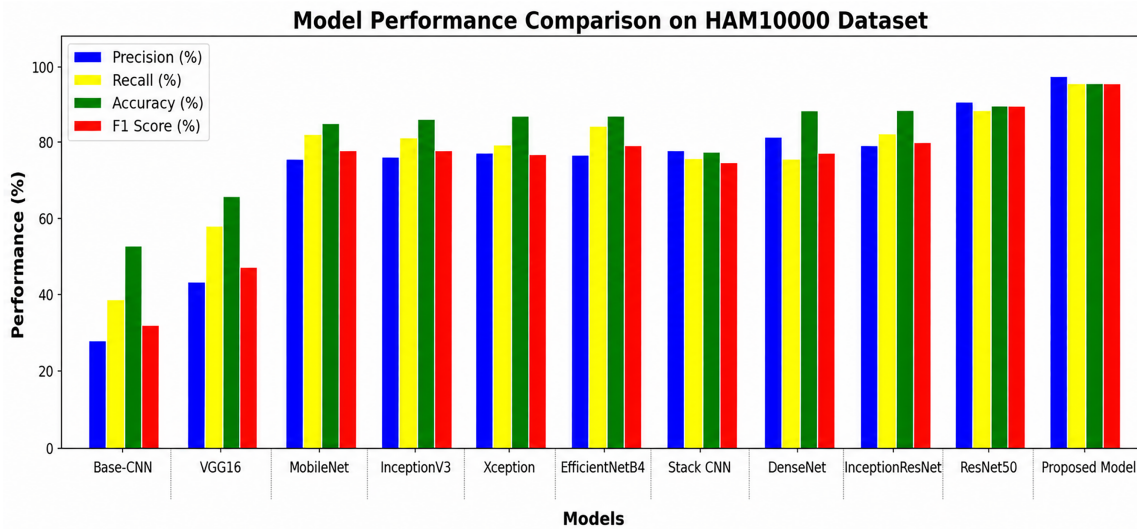


Figure 9: Classification accuracy of different models including the proposed model trained on HAM10000.

The training phase could be observed from the loss-epoch curve. Reducing the model's loss generally improves its predictive accuracy. Here, each model was trained for 50 epochs to estimate the error between predictions and actual outcomes. Fig. 10 shows the loss-epoch curves for both train loss (blue line) and validation loss (orange line) across different models. From the graph, it is clear that Base-CNN and VGG16 are the poorest performers. The training loss for Base-CNN remains almost unchanged over time, while VGG16 training loss initially reduces after which it remains the same. MobileNet, InceptionV3, Xception, and EfficientNetB4 give decent performance as their loss value decreases over the time. During training DenseNet, InceptionResNet, ResNet-50, and the proposed solution loss decreases gracefully up to approximately 10 epochs and converges at 40. From Table 3, it can be observed that ResNet-50 performed well on HAM10000 dataset.

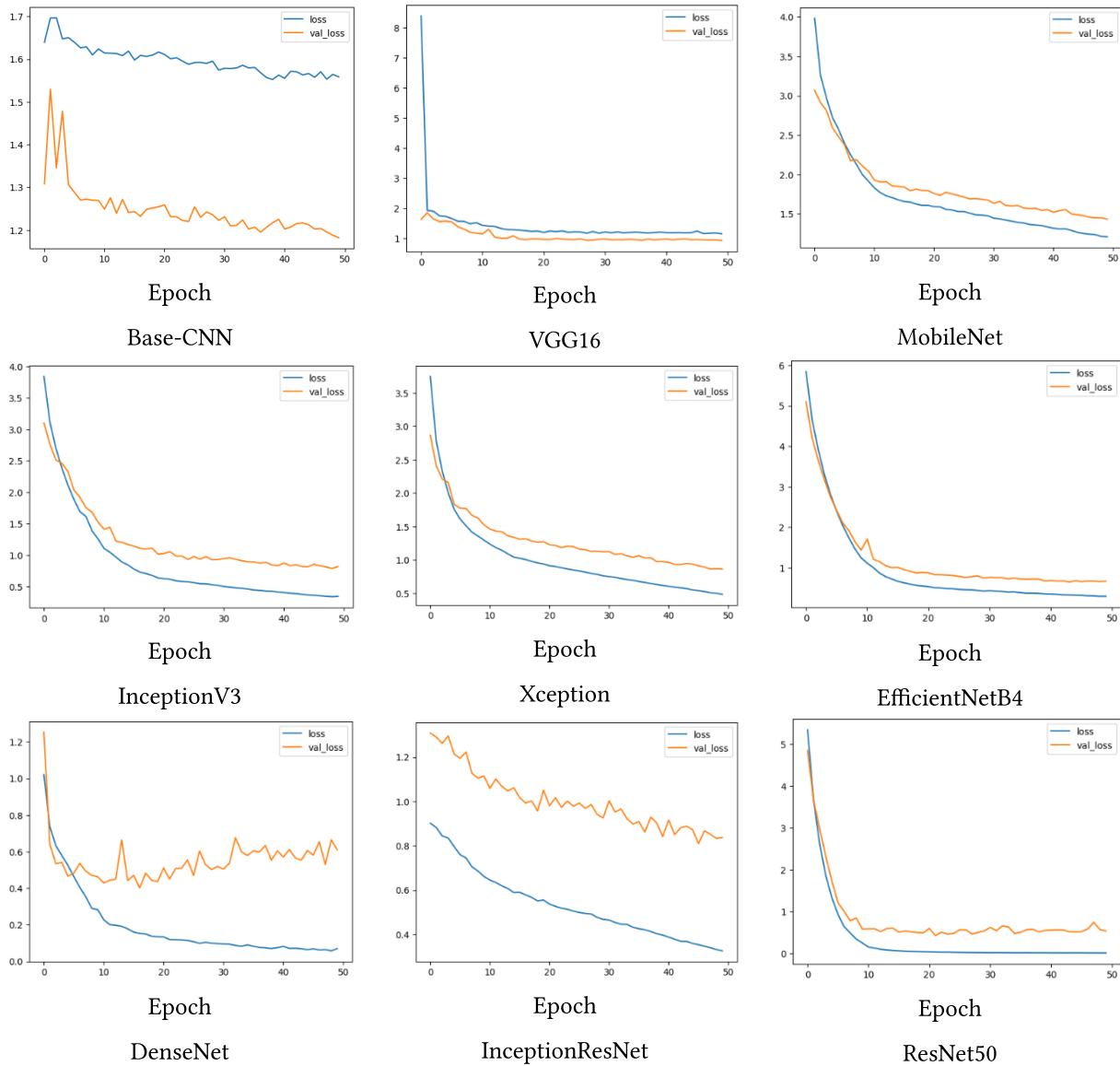
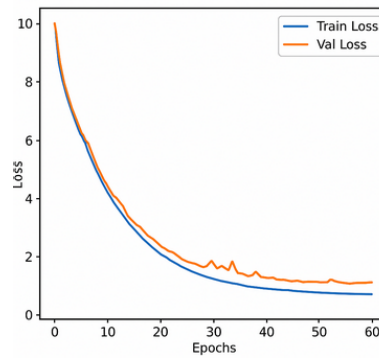


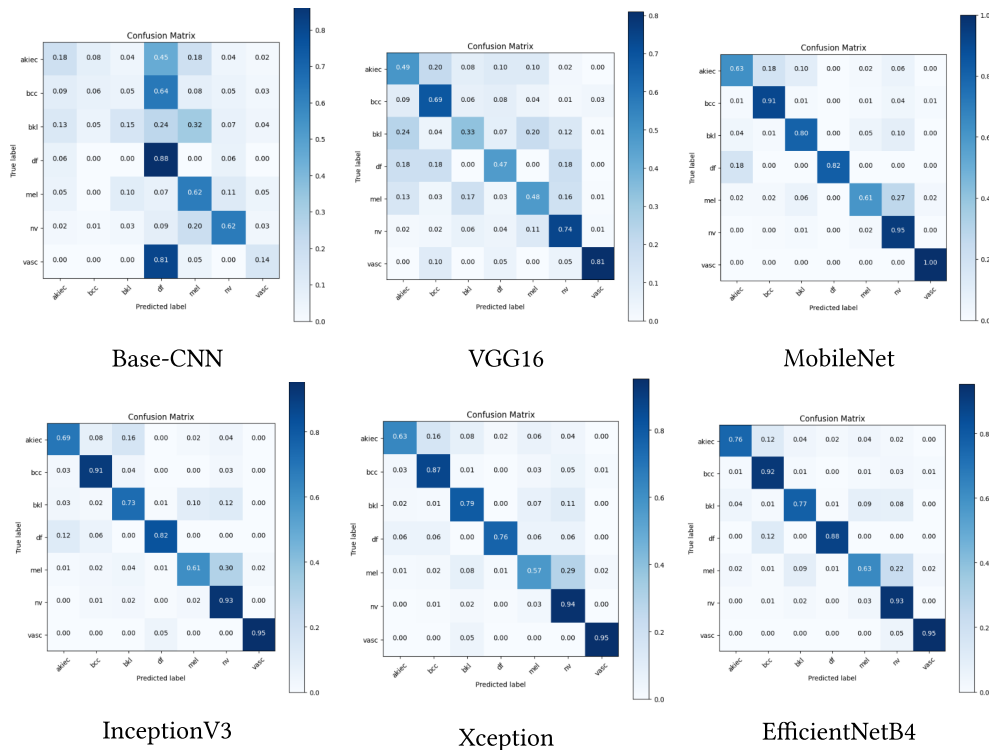
Figure 10: (Continued)



Proposed Solution

Figure 10: Loss curves of proposed model and baseline models trained on HAM10000 dataset for different models.

The confusion matrix gives an overview on the model's performance of the different classes by relating the predicted labels to the actual labels in a tabular form. The Fig. 11 pictorial representation of the performance of HAM10000 dataset, which contains seven classes with predicted labels is shown on x -axis and true labels on y -axis. From the confusion matrix, Base-CNN is the worst performer, as it can be seen that multiple classes are below 20% accuracy. VGG16 shows little improvement, with number of classes below 50% accuracy. Xception, DenseNet and ResNet50 show further improvement in accuracy, whereas InceptionV3, MobileNet, EfficientNet and InceptionResnet beat the above models, each achieving over 90% accuracy in three classes on HAM10000 dataset. But the proposed solution outperforms by achieving 97% accuracy. The proposed solution is able to provide better accuracy for generalization across classes.

**Figure 11:** (Continued)

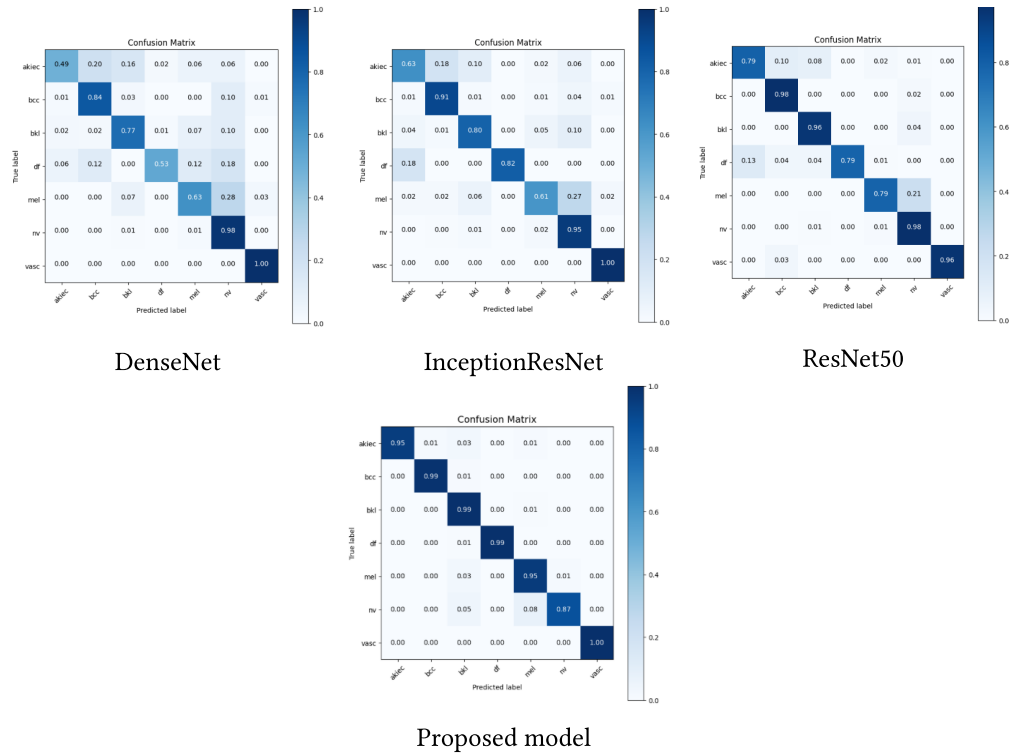


Figure 11: Confusion matrix comparison of baseline models with proposed model trained on HAM10000 dataset.

The Receiver Operating Characteristic (ROC) curve in Fig. 12 shows how the proposed solution is able to distinguish distinct classes with great accuracy across seven individual groups of diagnosis. Multiple classes show AUC scores above or equal to 0.99, which indicates the model maintains high accuracy and prevents wrong positive predictions across generalizations. The initial rapid slope of the curve shows reliable model performance because it indicates both a high rate of true positives while maintaining a low rate of false positives.

The performance of the proposed solution is evaluated over four datasets, namely ISIC-2019-2020, ISIC, ISIC-2019, and HAM1000. A statistical significance test was conducted. Table 4 presents the mean and standard deviation of all five evaluation metrics across 10-fold cross-validation for all four datasets. The proposed framework demonstrates consistently low standard deviation across all metrics – $\pm 0.067\%$ for HAM10000, $\pm 0.205\%$ for ISIC-2019-2020, $\pm 0.392\%$ for ISIC, and $\pm 0.497\%$ for ISIC-2019.

For binary classification, the model performance is examined on ISIC-2019-2020 dataset, where it achieves F1-score of 93.81%, accuracy of 93.91%, and area under the curve of 97.63% confirming high confidence for differentiating mel from benign lesions. The same experiment was conducted on the ISIC dataset with the model F1-score of 93.48%, accuracy of 93.53%, and the AUC score of 98.27%, showing its generalization across the datasets. On the ISIC-2019 dataset with eight distinct classes, the proposed solution shows its ability to handle complex multi-class classification tasks. The training accuracy reaches 99% while validation accuracy remains close to 91.47%, and an AUC of 98.17% indicates good generalization.

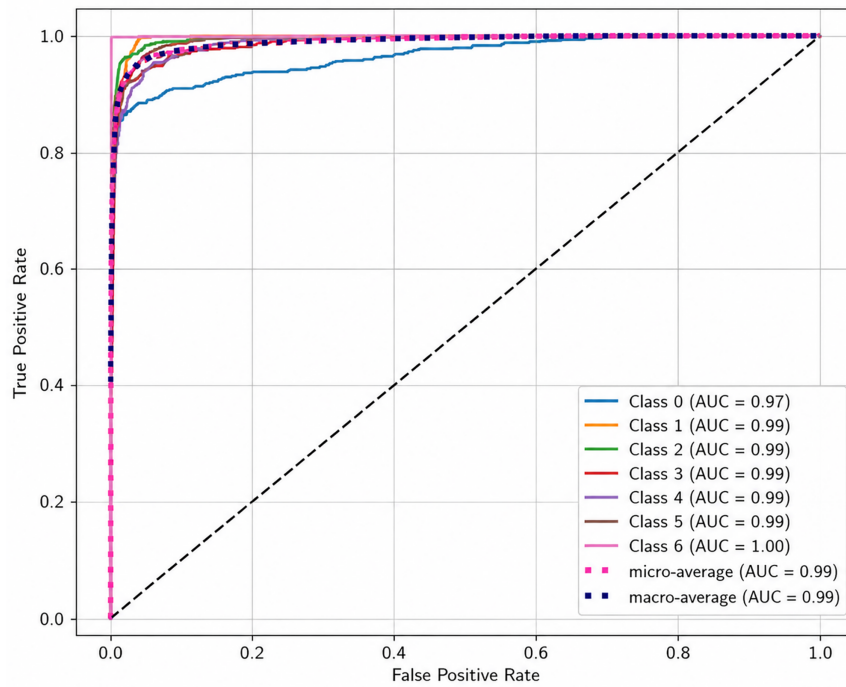


Figure 12: ROC curve of proposed solution trained using HAM10000 dataset.

Table 4: Statistical analysis of the proposed model across 10-fold cross-validation on all datasets.

Model	Dataset	Class	Metric	Mean (%)	Std (%)
Proposed	ISIC-2019-2020	2	Accuracy	93.91	0.205
			Precision	94.02	0.202
			Recall	93.66	0.215
			F1-Score	93.81	0.211
			AUC	97.63	0.144
Proposed	ISIC	2	Accuracy	93.53	0.392
			Precision	93.45	0.406
			Recall	93.51	0.394
			F1-Score	93.48	0.397
			AUC	98.27	0.208
Proposed	ISIC-2019	8	Accuracy	91.47	0.497
			Precision	91.86	0.555
			Recall	91.33	0.559
			F1-Score	91.51	0.574
			AUC	98.17	0.094
Proposed	HAM10000	7	Accuracy	97.77	0.067
			Precision	98.00	0.061
			Recall	97.17	0.059
			F1-Score	97.45	0.064
			AUC	99.00	0.009

The proposed solution shows the training loss and validation loss in Fig. 13 across three different datasets. The graph demonstrates models learning ability and generalization capability. For ISIC-2019-2020 dataset, the training loss and validation loss are decreasing consistently with the increase in epochs. This indicates that the model is learning successfully. In the ISIC dataset, both training and validation loss continuously reduce during the training period. This indicates that the model demonstrates good generalization ability and leads to enhanced learning performance. For the ISIC-2019 dataset, there is a decrease in training loss during the initial phase, eventually reaching a steady state. In the beginning, the validation loss reduced, then maintained a steady pattern, which suggests convergence occurred effectively. This suggests that the model is able to extract the important features from the dataset and able to generalize.

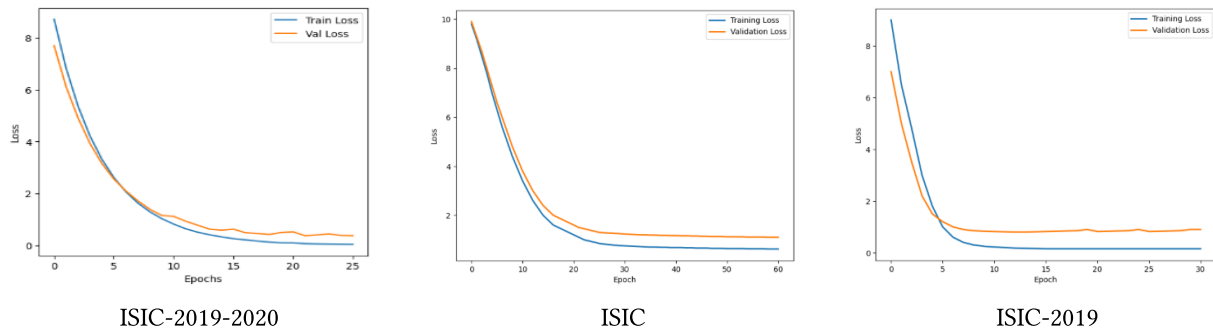


Figure 13: Loss-epoch curve on ISIC datasets.

Fig. 14 shows the confusion matrices of the proposed solution for three distinct datasets. It can be observed through the matrices how well the proposed solution separates different classes. The proposed solution shows excellent classification capability when differentiating nevus and melanoma within the ISIC-2019-2020 data. The classification success rate for nevus detection acquires 93% while the melanoma reaches to 96%. For ISIC dataset, the model is able to separate benign and malignant diagnosis. The model achieves 94% accuracy for benign class and 93% for malignant class. The ISIC-2019 dataset consists of eight types of skin lesions. According to the confusion matrices shown in Fig. 14, the proposed model has achieved precision for DF and VASC (100%), followed by SCC (99.01%), AK (95.10%), BCC (91.84%), and BKL (91.40%). The MEL and NV classes obtained precision of 86.84% and 74.19%, respectively.

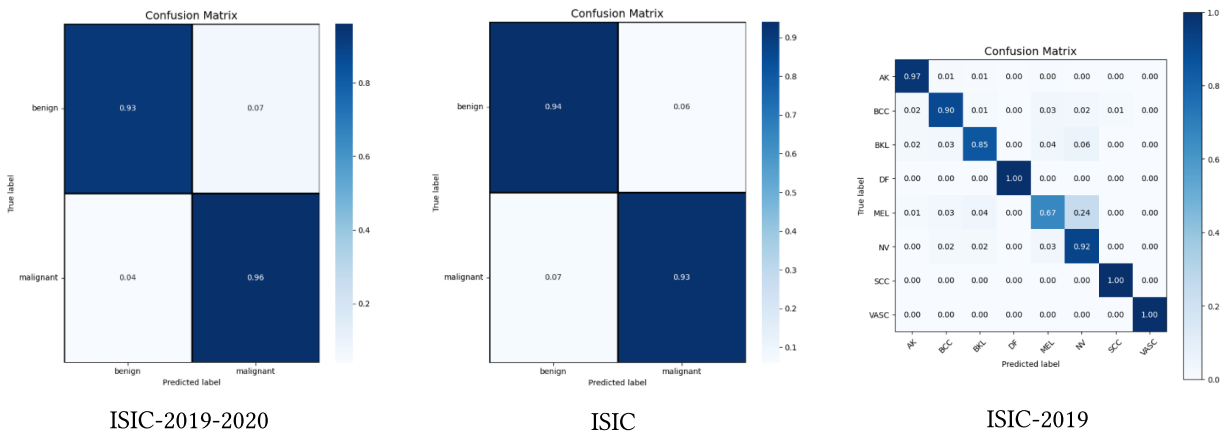


Figure 14: Confusion metrics on ISIC datasets.

The receiver operating characteristic (ROC) curve is applied on three dataset ISIC-2019-2020, ISIC, and ISIC-2019 displayed in Fig. 15. Our solution performance is demonstrated through ROC curve, which evaluates the true positive rate (TPR) and false positive rate (FPR) to determine the area under the curve (ROC-AUC), which leads to performance indicator. The ISIC-2019-2020 dataset shows an AUC score of 97.63%, which indicates model's high capability in distinct classes. The ISIC dataset show an AUC score of 98.27%. From this curve, it can be observed that there is a rise in TPR while FPR remains low due to strong ability to distinguish benign and malignant images. The individual AUC score for ISIC-2019 dataset of different classes, ranging from 95% to 100%, indicates that the features are perfectly separated as shown in Fig. 15.

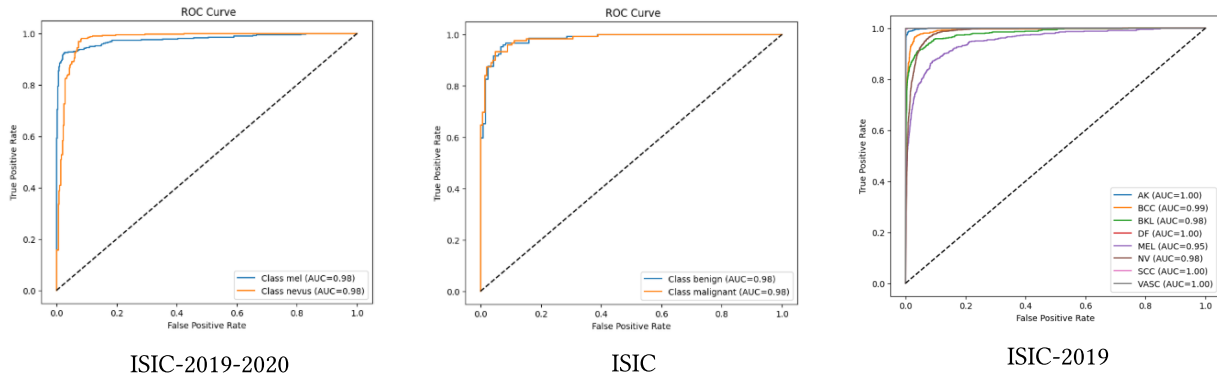


Figure 15: ROC curve on ISIC datasets.

Table 5: Comparison of previous research works with the proposed model on different datasets.

State of the Art Models	Year	Dataset	Class	Accuracy	AUC
Attention-Based CNN [38]	2025	HAM10000	7	86.83	86.20
Transfer Learning with Zero Shot [18]	2025	HAM10000	7	93.00	87.00
CNN-DenseNet [50]	2024	HAM10000	7	93.24	94.06
Lightweight ViT [19]	2024	HAM10000	7	95.82	93.00
Ensemble Deep Network [20]	2024	HAM10000	7	95.87	–
DEnSha [21]	2023	HAM10000	7	96.20	–
Proposed		HAM10000	7	97.77	99.00
Dermoscopic Augmented CNN Framework [22]	2024	ISIC-2019	8	88.63	94.00
Fused CNN PCA Classifier [23]	2023	ISIC-2019	8	96.10	94.41
Proposed		ISIC-2019	8	91.47	98.17
Fine-Tuned Transfer Learning Models [24]	2024	ISIC	2	93.50	86.10
Hybrid InceptionV3–DenseNet121 Model [25]	2024	ISIC	2	92.27	–
Proposed		ISIC	2	93.53	98.27
Weighted Ensemble of CNNs [15]	2023	ISIC-2019-2020	2	93.36	91.07
Layer Optimized CNN Framework [51]	2024	ISIC-2019-2020	2	90.85	90.07
Proposed		ISIC-2019-2020	2	93.91	97.63

Table 5 shows the performance of previous research works on HAM10000 and ISIC datasets. Compared to other models proposed by previous researches [15,18,22–24,50,52], the proposed solution performs better in terms of AUC score for ISIC-2019 (98.17), ISIC(98.27), and ISIC-2019-2020 (97.63). We have shown in **Table 5** a comparison between the proposed framework and several existing methods across multiple benchmark datasets. On the HAM10000 dataset, the proposed model achieved an accuracy of 97.77% and an AUC of 99.00%, surpassing the transfer and zero-shot learning approach of Chowdhury et al. [18] and the hybrid CNN model of Khan et al. [50]. For the ISIC-2019 dataset, our method attained an AUC of 98.17%, outperforming Nugroho et al. [22] and Olayah et al. [23], who reported AUCs of 94.00% and 94.41%, respectively. On the ISIC binary classification dataset, the proposed solution reached an accuracy of 93.53% and an AUC of 98.27%, exceeding the performances reported by Islam and Panta [24] and Akter et al. [25]. Finally, for the ISIC-2019-2020 dataset, our approach achieved an accuracy of 93.91% and an AUC of 97.63%, outperforming the ensemble models of Meswal et al. [15] and Mahmoud et al. [51].

4.2 Ablation Study

We have conducted an ablation study on the HAM10000 dataset by adding each module to the ResNet-50 baseline. As shown in **Table 6**, the baseline ResNet-50 without any additional components achieves an accuracy of 90.00% and F1-score of 90.00%. Adding dropout improves accuracy to 91.20% by reducing overfitting. The inclusion of L2-regularization further improves accuracy to 92.10% by penalizing large weights and improving generalization. The addition of the self-attention module produces a significant jump to 94.50%, confirming that attention-guided feature learning effectively focuses on diagnostically relevant lesion regions. The dual pooling strategy yields a notable improvement to 95.30%. Finally, PSO-based hyperparameter optimization achieves the highest accuracy of 97.77% and AUC of 99.00%, demonstrating that automated hyperparameter tuning is critical for maximizing model performance. These results show that each component plays an important role in improving the overall performance of the proposed framework. Furthermore, the proposed framework is compared to recent attention-based skin cancer classification techniques in **Table 7**. As can be seen from the comparison, the proposed Hybrid Attention-Enhanced ResNet-50 model outperforms the existing attention-based models, suggesting that the combination of self-attention with the proposed training and optimization strategy is a more powerful feature-learning framework than those of the existing attention-based models.

Table 6: Ablation study on HAM10000 dataset showing contribution of each component.

Model Configuration	Self-Attention	Dual Pooling	PSO	L2 Reg	Drop Out	Acc (%)	F1 (%)	AUC
ResNet-50 baseline	✗	✗	✗	✗	✗	90.00	90.00	95.60
+ Dropout	✗	✗	✗	✗	✓	91.20	91.10	96.10
+ L2 Regularization	✗	✗	✗	✓	✓	92.10	91.90	96.50
+ Self-Attention	✓	✗	✗	✓	✓	94.50	94.20	97.10
+ Dual Pooling	✓	✓	✗	✓	✓	95.30	95.10	98.40
+ PSO (Full Model)	✓	✓	✓	✓	✓	97.77	97.45	99.00

Table 7: Comparison of attention-based methods for skin cancer classification.

Authors	Year	Methodology	Performance (%)
Shu et al. [53]	2024	Attention-based ConvNeXt	Acc = 77.60
Tai et al. [54]	2024	Self attention-based DC-AC	AUC = 90.00
Younas et al. [55]	2025	Inception-Residual CNN + Soft Attention	Acc = 91.63
Alotaibi and AlSaeed [56]	2025	Xception + Self-/Soft-/Hard-Attention modules	Acc = 94.11
Yasir and Kim [57]	2025	SE-based depthwise separable convolution	Acc = 94.41
Proposed	2025	Hybrid Attention-Enhanced ResNet-50 model	Acc = 97.77

4.3 Computational Complexity Analysis

Table 8 presents the computational analysis of the proposed framework. The proposed model contains 24,116,107 total parameters (92.00 MB), of which 24,062,987 (91.79 MB) are trainable, and 53,120 (207.50 KB) are non-trainable. The marginal increase in parameters compared to the standard ResNet-50 backbone is attributed to the additional self-attention module and dual pooling layers, which together contribute to improved feature representation. All experiments were conducted on a Tesla V100-SXM2-32GB GPU. Training time, inference time, and memory requirements are reported for the proposed model and all baseline models.

Table 8: Computational complexity analysis of the proposed framework and baseline models on HAM10000 dataset.

Model	Total Params (M)	Trainable Params (M)	Training Time (min)	Inference Time (ms)	GPU Memory (GB)
ResNet-50	23.6	23.6	30	8.2	4.2
VGG16	138.4	138.4	78	12.4	7.8
MobileNet	4.2	4.2	22	3.1	2.1
InceptionV3	23.9	23.9	52	9.8	4.8
DenseNet	28.7	28.7	68	11.2	5.9
Proposed	24.1	24.0	62	9.6	5.1

4.4 Visual Interpretation and Explainability

Fig. 16 presents feature map visualization using Grad-CAM for the proposed Hybrid Attention-Enhanced ResNet-50 model. Grad-CAM highlights the regions of the predicted image that are most significant for classification. Each column corresponds to a distinct skin lesion category from the HAM10000 dataset and depicts three progressive stages: the original dermoscopic image (top), the baseline Grad-CAM visualization derived from the standard ResNet-50 model (middle), and the refined Grad-CAM heatmap generated after integrating the hybrid self-attention mechanism (bottom).

The Grad-CAM shows the regions most important to the ResNet-50 model. In some cases, these areas are more extensive than the lesion itself and involve skin around the lesion. After applying the attention mechanism, the visualizations are much clearer and more accurate. Now the refined Grad-CAM maps are more focused on the actual boundaries and internal structures of the lesion and provide a more general understanding of what's important in the image. In Malignant cases like MEL and BCC, the attention enhanced maps highlight clearly the irregular margins and the regions with different pigmentation. For non-cancerous lesions such as DF and NV, the attention mechanism suppresses background noise and draws its attention to the central part of the lesion, emphasizing the smooth edges and even color typical of benign skin growths.

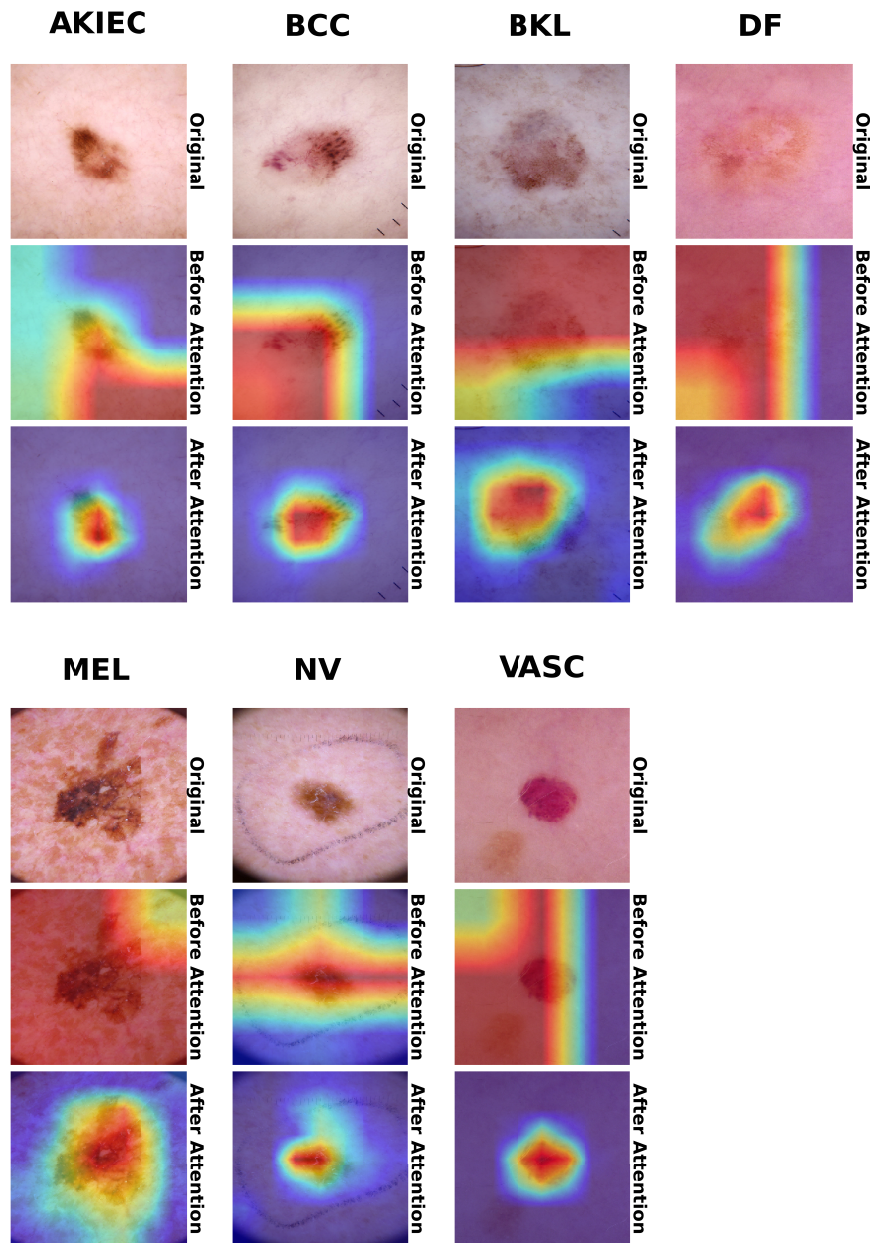


Figure 16: Comparison of feature vectors extracted (using Grad-CAM) from HAM10000 dataset images across seven skin lesion classes. Each column represents a distinct class. The figure illustrates the original image samples, the feature vectors obtained from ResNet-50, and the refined feature vector representations after applying the self-attention mechanism.

This clearer visualization demonstrates the ability of the hybrid attention module to be able to pay attention to the most important and clinically relevant components of the lesion, while suppressing irrelevant information from the surrounding skin.

4.5 Discussion

We integrate transfer learning, self-attention, and dual pooling in the Hybrid Attention-Enhanced ResNet-50. We use the pre-trained ResNet-50 model as a feature extraction backbone. The self-attention

mechanism enhances the representation of diagnostically relevant spatial relationships by reducing the influence of irrelevant background regions. This contribution is supported by the ablation results in Table 6, where the ResNet-50 baseline achieves 90.00% accuracy and the addition of self-attention increases the accuracy to 94.50%, and the F1-score to 94.20%. The subsequent integration of dual pooling further improves the accuracy to 95.30%. It indicates that combining dominant local responses with global contextual information provides richer representations of skin lesions. The overall model achieves 97.77% accuracy, 97.45% F1-score, and 99.00% AUC after PSO-based optimization. This progressive improvement supports the contribution of combining attention-guided feature learning with complementary pooling and regularization strategies rather than relying on the ResNet-50 backbone.

We use PSO for the training optimization and the hyperparameter tuning. As shown in Table 6, dropout and L2-regularization improve the baseline accuracy from 90.00% to 91.20% and 92.10%, respectively. It suggests that regularization helps control overfitting during training. The final application of PSO increases accuracy from 95.30% to 97.77% and from 98.40% to 99.00% AUC. It indicates that optimized learning and regularization parameters can improve the attention-enhanced architecture. The loss curves in Figs. 10 and 13 demonstrate steady decrease in training and validation loss, further confirming stable convergence. Additionally, the statistical results in Table 4 show the std obtained across 10-fold cross validation are relatively low, suggesting that the results were consistent for different splits of the data. Thus, the observed performance is supported by the combined training strategy. Also, the ablation results specifically demonstrate the incremental contribution of dropout, L2-regularization, self-attention, dual pooling, and PSO.

The cross-dataset experiments provide evidence for the robustness and transferability of the proposed framework across both binary and multi-class skin lesion classification tasks. As reported in Table 4, the model achieves 93.91% accuracy and 97.63% AUC on ISIC-2019-2020, 93.53% accuracy and 98.27% AUC on ISIC, and 91.47% accuracy and 98.17% AUC on ISIC-2019. The confusion matrices in Fig. 14 show strong class-level discrimination, although the MEL and NV classes in ISIC-2019 remain comparatively more challenging, with precision values of 86.84% and 74.19%, respectively. The ROC curves in Fig. 15 further demonstrate strong discriminative capability, with AUC values of 97.63% for ISIC-2019-2020 and 98.27% for ISIC, while the individual ISIC-2019 class AUC values range from 95.00% to 100.00%. Comparison with previous studies in Table 5 shows that the proposed framework achieves higher AUC values on the evaluated ISIC datasets and strong performance on HAM10000. However, on ISIC-2019, the proposed model has a lower accuracy than the Fused CNN PCA Classifier reported in the literature, while achieving a higher AUC. Thus, the results indicate that the proposed framework provides strong discrimination and generalization, rather than uniformly outperforming every existing method on every metric.

The explainability analysis shows evidence for the contribution of self-attention. Original dermoscopic images are compared in Fig. 16 with the Grad-CAM representation derived from the baseline ResNet-50 and the refined representation obtained by adding self-attention. The attention-enhanced maps show a more focused response around lesion boundaries and internal structures. It reduces activation in surrounding skin regions. This behavior is particularly visible for MEL and BCC samples, where the enhanced maps emphasize irregular margins and pigmentation patterns, while DF and NV samples show more concentrated activation within the lesion region. These visual observations complement the quantitative improvement observed in the ablation study and provide evidence that the attention mechanism changes the feature representation in a diagnostically meaningful direction. The computational analysis in Table 8 also shows that this improvement involves a modest increase in model complexity, with 24.1M parameters and 9.6 ms inference time compared with 23.6M parameters and 8.2 ms for the standard ResNet-50.

5 Conclusion

To address the challenges associated with feature extraction arising from the intricate nature of human skin, which hinders automated classification, we proposed a framework based on the ResNet-50 backbone with transfer learning, a self-attention mechanism, hybrid pooling (max pooling and global average pooling), and several regularization and optimization techniques. We incorporated L2-regularization, dropout, He normal initialization, adaptive learning rate scheduling, 10-fold cross-validation, and early stopping to improve model stability and reduce overfitting. Furthermore, we fine-tuned the model hyperparameters using PSO to achieve more stable convergence. In addition, we employed the explainable AI technique Grad-CAM to improve model interpretability by highlighting clinically relevant regions, thereby increasing the transparency and reliability of the classification results.

The model obtained the maximum accuracy of 97.77% and AUC of 99.00% on HAM10000 dataset with seven classes. The proposed framework showed superior classification performance on the three different ISIC datasets, with accuracy and AUC scores of 93.91%, 93.53%, and 91.47% and 97.63%, 98.27%, and 98.17%, respectively.

The results obtained from the proposed model are compared with the results reported in the available literature. The comparison shows that the proposed method outperforms in terms of classification performance with respect to different evaluation metrics.

In future, we will work on generating missing data and integrating automated hair removal techniques into this proposed framework. Moreover, Vision Transformers (ViTs), including Swin Transformer and DeiT, attention modules such as the Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation (SE) attention, and Efficient Channel Attention (ECA), as well as ResNet101, DenseNet201, EfficientNetB4, and ConvNeXt, will be explored. We will focus on improving feature extraction, class separation, and the reliability of diagnosis for an automated skin cancer detection system.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Chanchal Biswas: Conceptualization, methodology, data curation, validation, writing—original draft preparation. Naveenkumar Mahamkali: Supervision, reviewing and editing. Satish Anamalamudi, Bhaskar Marapelli, Ch Anil Carie: Reviewing and editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: All datasets used in this research are publicly available online.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yasmin J, Sathik M. An improved iterative segmentation algorithm using canny edge detector for skin lesion border detection. *Int Arab J Inf Technol (IAJIT)*. 2015;12(4):37–42. doi:10.5120/7779-0865.
2. Zortea M, Skrovseth SO, Schopf TR, Kirchesch HM, Godtliebsen F. Automatic segmentation of dermoscopic images by iterative classification. *Int J Biomed Imaging*. 2011;2011(1):972648. doi:10.1155/2011/972648.
3. Khattar S, Kaur R. Computer assisted diagnosis of skin cancer: a survey and future recommendations. *Comput Electr Eng*. 2022;104:108431.
4. Marosán-Vilimszky P, Szalai K, Horváth A, Csabai D, Füzesi K, Csány G, et al. Automated skin lesion classification on ultrasound images. *Diagnostics*. 2021;11(7):1207. doi:10.3390/diagnostics11071207.

5. Ahmad N, Shah JH, Khan MA, Baili J, Ansari GJ, Tariq U, et al. A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI. *Front Oncol.* 2023;13:1151257. doi:10.3389/fonc.2023.1151257.
6. Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S. Skin cancer classification with deep learning: a systematic review. *Front Oncol.* 2022;12:893972. doi:10.3389/fonc.2022.893972.
7. Saba T, Khan MA, Rehman A, Marie-Sainte SL. Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction. *J Med Syst.* 2019;43(9):289.
8. Khan ZA, Waqar M, Khan HU, Chaudhary NI, Khan AT, Ishtiaq I, et al. Fine-tuned deep transfer learning: an effective strategy for the accurate chronic kidney disease classification. *PeerJ Comput Sci.* 2025;11:e2800.
9. Kumar R, Kumbharkar P, Vanam S, Sharma S. Medical images classification using deep learning: a survey. *Multimed Tools Appl.* 2024;83(7):19683–728. doi:10.1007/s11042-023-15576-7.
10. Alrabai A, Echtioui A, Kallel F. Exploring pre-trained models for skin cancer classification. *Appl Syst Innov.* 2025;8(2):35. doi:10.3390/asi8020035.
11. Ahmed KT, Rustam F, Mehmood A, Ashraf I, Choi GS. Predicting skin cancer melanoma using stacked convolutional neural networks model. *Multimed Tools Appl.* 2024;83(4):9503–22. doi:10.1007/s11042-023-15488-6.
12. Serener A, Serte S. Keratinocyte carcinoma detection via convolutional neural networks. In: *Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); 2019 Oct 11–13; Ankara, Turkey.* p. 1–5.
13. Nivyashree T, Pramila P. Detection of malignant and benign skin lesions using the influence of activation function and accuracy analysis in densely connected convolutional network compared over convolutional neural network. In: *Proceedings of the 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS); 2023 Dec 14–15; Chennai, India.* p. 1–6.
14. Kollipara VH, Kollipara VDP. Residual learning based approach for multi-class classification of skin lesion using deep convolutional neural network. In: *International Conference on Cognition and Recongition.* Cham, Switzerland: Springer; 2021. p. 340–51.
15. Meswal H, Kumar D, Gupta A, Roy S. A weighted ensemble transfer learning approach for melanoma classification from skin lesion images. *Multimed Tools Appl.* 2024;83(11):33615–37. doi:10.1007/s11042-023-16783-y.
16. Akilandasowmya G, Nirmaladevi G, Suganthi S, Aishwariya A. Skin cancer diagnosis: leveraging deep hidden features and ensemble classifiers for early detection and classification. *Biomed Signal Process Control.* 2024;88:105306.
17. Maurya R, Bais AK, Gopalakrishnan T, Dutta MK, Pandey NN, YV SM. Skin lesion classification using deep feature fusion and selection using XGBoost classifier. In: *Proceedings of the 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS); 2024 Feb 24–25; Bhopal, India.* p. 1–5.
18. Chowdhury TA, Wagner E, Motzki P, Lehser M. Enhanced transfer learning algorithm with zero-shot components for dermatological diagnosis using the HAM10000 dataset. In: *Proceedings of the Photonics in Dermatology and Plastic Surgery 2025; 2025 Jan 25–27; San Francisco, CA, USA.* Vol. 13292. SPIE. p. 116–23.
19. Dwivedi T, Chaurasia BK, Shukla MM. Lightweight vision image transformer (LViT) model for skin cancer disease classification. *Int J Syst Assur Eng Manag.* 2024;15(10):5030–55. doi:10.1007/s13198-024-02521-6.
20. Srivastava V, Gupta S, Singh R, Gautam V. A multi-level closing based segmentation framework for dermoscopic images using ensemble deep network. *Int J Syst Assur Eng Manag.* 2024;15(8):3926–39. doi:10.1007/s13198-024-02393-w.
21. Arora G, Dubey AK, Jaffery ZA. Multiple skin lesion classification using deep, ensemble, and shallow (DEnSha) neural networks approach. *Int J Syst Assur Eng Manag.* 2023;14(Suppl 1): 385–93. doi:10.1007/s13198-023-01866-8.
22. Nugroho ES, Ardiyanto I, Nugroho HA. Addressing imbalance ISIC-2019 dataset in dermoscopic pigmented skin lesion classification. *ICIC Express Lett.* 2024;18(6):563–73.
23. Olayah F, Senan EM, Ahmed IA, Awaji B. AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features. *Diagnostics.* 2023;13(7):1314. doi:10.3390/diagnostics13071314.

24. Islam MS, Panta S. Skin cancer images classification using transfer learning techniques. arXiv:2406.12954. 2024. doi:10.21203/rs.3.rs-6473686/v1.
25. Akter M, Khatun R, Talukder MA, Islam MM, Uddin MA, Ahamed MKU, et al. An integrated deep learning model for skin cancer detection using hybrid feature fusion technique. *Biomed Mater Devices*. 2025;3(2):1433–47. doi:10.1007/s44174-024-00264-3.
26. Ingle YS, Shaikh NF. Skin cancer recognition using attention-based CNN algorithm. In: *Proceedings of the 2024 IEEE Pune Section International Conference (PuneCon)*; 2024 Dec 13–15; Pune, India. p. 1–10.
27. Yao Y, Lilhore UK, Simaiya S, Aldossary SM, Tekeste LG, Ghith E, et al. Advancing skin cancer diagnosis with deep learning and attention mechanisms. *Sci Rep*. 2026;16:3610. doi:10.1038/s41598-025-33597-0.
28. Lilhore UK, Sharma YK, Simaiya S, Alroobaea R, Baqasah AM, Alsafyani M, et al. SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems. *Sci Rep*. 2025;15(1):14913. doi:10.1038/s41598-025-98205-7.
29. Saeed MA, Afify YM, Badr NL, Helal NA. Multimodal deep learning ensemble framework for skin cancer detection. *Sci Rep*. 2025;15(1):45660. doi:10.1038/s41598-025-30534-z.
30. Islam S, Wishart GC, Walls J, Hall P, de Herrera GS, Gan A, et al. Advancing skin cancer detection through deep learning and fusion of patient metadata and skin lesion images. *Sci Rep*. 2026;16(1):1968.
31. Vieira J, Mendonça F, Morgado-Dias F. Deep learning approaches for skin lesion detection. *Electronics*. 2025;14(14):2785. doi:10.3390/electronics14142785.
32. Sethanan K, Pitakaso R, Srichok T, Khonjun S, Thannipat P, Wanram S, et al. Double AMIS-ensemble deep learning for skin cancer classification. *Expert Syst Appl*. 2023;234(7):121047. doi:10.1016/j.eswa.2023.121047.
33. Thepade SD, Shukla S. Enhancing melanoma skin cancer detection through feature fusion of pre-trained deep convolutional neural network ResNet50 and thepade sorted block truncation coding. *SN Comput Sci*. 2024;5(4):426.
34. Zhao H, Jia J, Koltun V. Exploring self-attention for image recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020 Jun 13–19; Seattle, WA, USA. p. 10076–85.
35. Wu Y, Ma Y, Liu J, Du J, Xing L. Self-attention convolutional neural network for improved MR image reconstruction. *Inf Sci*. 2019;490(6):317–28. doi:10.1016/j.ins.2019.03.080.
36. Khan SUR, Asif S. Oral cancer detection using feature-level fusion and novel self-attention mechanisms. *Biomed Signal Process Control*. 2024;95(1):106437. doi:10.1016/j.bspc.2024.106437.
37. Pedro R, Oliveira AL. Assessing the impact of attention and self-attention mechanisms on the classification of skin lesions. In: *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*; 2022 Jul 18–23; Padua, Italy. p. 1–8.
38. Hosny M, Elgendy IA, Chelloug SA, Albashrawi MA. Attention-based convolutional neural network model for skin cancer classification. *IEEE Access*. 2025;13:172027–50. doi:10.1109/access.2025.3616022.
39. Kumari A, Rattan P, Shukla AK, Rani S, Kataria A, Min H, et al. Deep learning-assisted early detection of skin cancer from dermoscopic images in underserved clinical settings. *Bioengineering*. 2026;13(4):456. doi:10.3390/bioengineering13040456.
40. Gabani V, Navamani T, Shyamala K, Vaswani Rajpal VK. Multimodal skin lesion classification for early cancer diagnosis using deep learning. *Front Physiol*. 2026;17:1717517. doi:10.3389/fphys.2026.1717517.
41. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;5(1):180161. doi:10.1038/sdata.2018.161.
42. Maharana P. ISIC 2019 skin lesion images for classification. 2022. Available from: <https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-imagesfor-classification>.
43. Deng Q. ISIC 2019 and 2020 melanoma dataset. 2025. Available from: <https://www.kaggle.com/datasets/qikangdeng/isic-2019-and-2020-melanoma-dataset/data>.
44. Fanconi C. Skin cancer: malignant vs. Benign, processed skin cancer pictures of the ISIC. 2025. Available from: <https://www.kaggle.com/fanconic/skin-cancer-malignant-vsbenign>.
45. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:6000–10. doi:10.65215/r5bs2d54.

46. Salian AC, Vaze S, Singh P, Shaikh GN, Chapaneri S, Jayaswal D. Skin lesion classification using deep learning architectures. In: Proceedings of the 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA); 2020 Apr 3–4; Mumbai, India. p. 168–73.
47. Innani S, Dutande P, Baheti B, Baid U, Talbar S. Deep learning based novel cascaded approach for skin lesion analysis. In: International Conference on Computer Vision and Image Processing. Cham, Switzerland: Springer; 2022. p. 615–26.
48. Huang HW, Hsu BWY, Lee CH, Tseng VS. Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers. *J Dermatol*. 2021;48(3):310–6. doi:10.1111/1346-8138.15683.
49. Fraiwan M, Faouri E. On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors*. 2022;22(13):4963. doi:10.3390/s22134963.
50. Khan AR, Mujahid M, Alamri FS, Saba T, Ayesha N. Early-stage melanoma cancer diagnosis framework for imbalanced data from dermoscopic images. *Microsc Res Tech*. 2025;88(3):797–809. doi:10.1002/jemt.24736.
51. Mahmoud H, Omer OA, Ragab S, Esmail H, Abdel-Nasser M. Classifying melanoma in ISIC dermoscopic images using efficient Convolutional Neural Networks and deep transfer learning. *Traitement Signal*. 2024;41(2):679. doi:10.18280/ts.410211.
52. Lama N, Stanley RJ, Nambisan A, Maurya A, Hagerty J, Stoecker WV. Increasing melanoma diagnostic confidence: forcing the convolutional network to learn from the lesion. *arXiv:2305.09542*. 2023.
53. Shu C, Yu L, Tian S, Shi X. MSMA: a multi-stage and multi-attention algorithm for the classification of multimodal skin lesions. *Biomed Signal Process Control*. 2024;93:106180.
54. Tai CEA, Janes E, Czarnecki C, Wong A. Double-condensing attention condenser: leveraging attention in deep learning to detect skin cancer from skin lesion images. *Sensors*. 2024;24(22):7231. doi:10.3390/s24227231.
55. Younas S, Sargano AB, You L, Habib Z. Attention-based inception-residual CNN: skin cancer diagnosis with attention-based inception-residual CNN model. *Information*. 2025;16(2):120.
56. Alotaibi A, AlSaeed D. Skin cancer detection using transfer learning and deep attention mechanisms. *Diagnostics*. 2025;15(1):99. doi:10.3390/diagnostics15010099.
57. Yasir SM, Kim H. Attention-based deep feature aggregation network for skin lesion classification. *Electronics*. 2025;14(12):2364. doi:10.3390/electronics14122364.