

Article

Not peer-reviewed version

Skin Cancer Disease Detection Using Two-Stream Hybrid Attention Based Deep Learning Model

[Abu Saleh Musa Miah](#)^{*}, Koki Hirooka, [Jungpil Shin](#)^{*}

Posted Date: 25 June 2025

doi: 10.20944/preprints202506.1913.v1

Keywords: computer-aided disease detection; skin cancer detection; medical imaging; attention; deep learning; transfer learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Skin Cancer Disease Detection Using Two-Stream Hybrid Attention Based Deep Learning Model

Abu Saleh Musa Miah *, Koki Hirooka and Jungpil Shin *

School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Japan

* Correspondence: musa@u-aizu.ac.jp, jpshin@u-aizu.ac.jp

Abstract

Skin cancer is a serious health issue that requires early detection and treatment for effective management. Timely and accurate evaluation of skin lesions is crucial, as delays can lead to more severe outcomes. However, identifying skin lesions accurately can be challenging due to differences in color, shape, and the various types of imaging equipment used for diagnosis. While recent studies have demonstrated the potential of ensemble convolutional neural networks (CNNs) for the early diagnosis of skin disorders, these models are often too large and inefficient for processing contextual information. Although lightweight networks like MobileNetV3 and EfficientNet have been developed to reduce parameters and enable deep neural networks on mobile devices, their performance is limited by inadequate feature representation depth. To mitigate these limitations, we propose a hybrid attention dual-stream deep learning model for skin lesion detection. With just one training loss, this model preprocesses the input images and uses two branches, each of which extracts local and global features using multi-stage and multi-branch attention techniques. The first branch processes the original image using a convolutional layer integrated with three novel attention modules: Enhanced Separable Depthwise Convolution (SCAttn), stage attention, and branch attention. The second branch enhances the input image using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast and reveal finer details, and then processes the CLAHE-enhanced image through a similar feature extraction pip. The integration of CLAHE with SCAttn modules enables detailed feature extraction while maintaining computational efficiency. A classification module receives the concatenated hierarchical features extracted from both branches. Utilizing the PAD2020 and ISIC 2019 datasets, we assessed the proposed model and obtained an accuracy rate of 98.59% on PAD2020, surpassing the state-of-the-art performance by 2%, with stable performance accuracy for the ISIC 2019 dataset. This illustrates the ability of the proposed model to integrate several attention mechanisms and feature enhancement methods, offering a reliable and effective approach for detecting skin cancer.

Keywords: computer-aided disease detection; skin cancer detection; medical imaging; attention; deep learning; transfer learning

1. Introduction

Skin cancer is one of the most common and significant types of cancer worldwide, representing approximately one-third of all cancer cases [1–10]. The rising incidence, especially in the United States, where over 9,500 [11] diagnoses are made daily, underscores the importance of early detection and intervention [12,13]. However, skin cancer diagnosis has historically been challenging, particularly due to variations in skin lesion characteristics such as color, shape, and the diverse types of diagnostic imaging systems used. Recent advances in imaging technology, including dermatoscopy, reflectance confocal microscopy (RCM), and 3D total body photography, have significantly improved the ability to analyze skin lesions, but traditional diagnostic approaches still rely heavily on subjective interpretations by healthcare providers, often resulting in variability and delays in diagnosis [14]. Although skin cancer is frequently curable, early detection and accurate diagnosis are crucial for achieving positive

treatment outcomes and improving patient survival rates [15–17]. Traditional approaches for diagnosing skin cancer and other diseases [18–20], like visual inspections and histopathological exams, often suffer from being time-consuming, subjective, and prone to high variability among different observers. Approximately 53% of individuals diagnosed with melanomas personally conduct a self-examination before seeking medical evaluation [21]. The evolution of cancer therapeutics has seen a remarkable shift from traditional treatments like surgery, radiation, and chemotherapy to modern therapies such as immunotherapy and targeted treatments. These approaches aim to either directly target cancer cells or modulate their surrounding microenvironment, reflecting the growing emphasis on precision medicine in cancer care [2]. Initially, surgery was the first rational approach to treating cancer prior to 1970, evolving with advancements in anesthesia, surgical techniques, and early cancer detection. Radiation therapy, introduced after the discovery of X-rays, became a key treatment modality for targeting tumors while minimizing damage to healthy tissue. From 1970 to 2023, treatments like tamoxifen for breast cancer and targeted therapies such as trastuzumab and imatinib revolutionized cancer care by focusing on specific genetic alterations, significantly improving survival rates. Looking ahead to future treatment, emerging therapies, including PROTACs, molecular glue degraders, and gene-targeted treatments, promise to offer more precise treatments with fewer side effects. Additionally, advancements in immunotherapy, antibody drug conjugates (ADCs), and oncolytic virus therapy provide exciting new avenues for more effective, personalized cancer treatment strategies [22]. The clinical skin examination is a highly accurate method for affirmative skin cancer screening [14]. The discussion on skin cancer treatment paradigms focuses on the debate between directly targeting cancer cells and modulating the tumor microenvironment (TME) to control tumor growth and metastasis [23]. The TME, which includes tumor, stromal, and immune cells, plays a critical role in tumor progression and immune evasion. Targeting stromal cells, such as cancer-associated fibroblasts (CAFs), has shown promise, as they contribute to tumor growth, metastasis, and chemotherapy resistance. However, direct approaches like genetic therapies and metabolic reprogramming remain vital for reducing tumor cell viability and enhancing treatment sensitivity. Despite these advances, tumor heterogeneity and genetic instability often lead to resistance. Therefore, combining strategies that target both tumor cells and the TME—along with immune modulation through checkpoint blockade—may offer a more effective, comprehensive treatment approach to improve patient outcomes [23].

Emerging trends in oncology are moving towards integrating traditional therapies with innovative approaches like immunotherapy, personalized cancer vaccines, and machine learning-assisted imaging techniques. These advancements are expected to enhance early detection and treatment accuracy, offering a more comprehensive understanding of cancer dynamics and improved therapeutic strategies. Key developments in cancer therapies include nanomedicines, stem cell therapy, ablation therapy, and proton beam therapy, all of which promise safer, more efficient treatment options with minimal invasiveness. Additionally, gene therapy and natural antioxidants are being explored for their potential in cancer prevention and treatment. While these emerging technologies offer great promise, patient safety and cost-benefit considerations remain essential. A multimodal approach combining surgery, radiotherapy, chemotherapy, and newer therapies tailored to individual patient needs will likely become the future standard of care, improving overall outcomes [24]. Clinical examinations require significant time from physicians to analyze multiple dermoscopic images. Delays of weeks or months can postpone treatment and worsen skin conditions. Advances in imaging technology now enable more accurate skin lesion diagnosis through techniques such as reflectance confocal microscopy (RCM), 3D total body photography (3D-TBP), teledermatology, and dermoscopy. Dermoscopy, a non-invasive method, uses magnification to study lesions without relying on reflected light [14]. However, manual image processing is tedious, and results depend on subjective interpretation, making accurate skin lesion identification challenging. Previous methods, such as histogram thresholding and Principal Component Analysis (PCA) for colour space clustering, attempted to improve segmentation [25,26]. While these approaches reduce human involvement, they still struggle to distinguish all relevant features of skin lesions. In line with these advances, the role of deep learning in skin cancer detec-

tion has evolved significantly. Early approaches, such as those utilizing histogram thresholding or gradient vector flow, were limited by the need for significant human involvement and the inability to capture complex features. However, Convolutional Neural Networks (CNNs) [27,28], particularly with transfer learning, have proven highly effective for automating skin lesion analysis, achieving impressive accuracy levels [15,29,30]. While deep learning models like MobileNetV3 and EfficientNet have been used to facilitate skin cancer detection on mobile devices, they often fall short in terms of feature representation depth, limiting their accuracy [31–34]. To improve the performance accuracy, Hossain et al. proposed a Max Voting Ensemble Technique combining multiple pre-trained CNNs (e.g., ResNet50, InceptionV3, Xception) to enhance classification performance. This approach achieved 93.18% accuracy and an AUC of 0.9320 on the ISIC 2018 dataset [35]. Despite its effectiveness, the ensemble model is computationally expensive, making it unsuitable for real-time or resource-constrained applications. Further efforts have focused on optimizing Depthwise Separable Convolutions (DWS-Conv) to reduce model size while preserving performance [36]. Recent research has demonstrated that lightweight attention based deep learning for direct skin cancer detection, even on clinical PCs and mobile devices with limited processing resources [31,37–40]. Some hybrid models combine CNNs and Vision Transformers (ViTs) using either standard or DWSCov layers to compress feature dimensions before processing with transformer modules [39,41–44]. Most recently, Dai et al. proposed HierAttn, a lightweight network that employs a multi-stage, multi-branch attention mechanism guided by a single deep supervision loss [45]. HierAttn achieved 96.70% accuracy on ISIC2019 and 91.22% on PAD-UFES-20, outperforming other lightweight models. It addresses the insufficient feature depth of earlier lightweight networks and the computational burden of heavier ensemble methods. To further enhance performance, we propose a Two-Stream Hybrid Attention-Based Deep Learning Model. The first branch processes the original image through a convolutional layer integrated with three attention modules: Enhanced Separable Depthwise Convolution (SCAttn), stage attention, and branch attention. The second branch applies Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance local contrast and fine details, then follows a similar feature extraction pipeline. The main contributions of the proposed model are given below:

- We employed a novel *hybrid attention dual-stream deep learning model* specifically designed for skin lesion detection, which addresses the limitations of existing ensemble CNN models that are often too large and inefficient for contextual information processing.
- Our model effectively utilizes two distinct feature extraction branches: the first branch integrates a convolutional layer with three novel attention modules—*Enhanced Separable Depthwise Convolution (SCAttn), branch attention, and stage attention*—to capture both local and global hierarchical features from the original images.
- We enhance the input images in the second branch using *Contrast Limited Adaptive Histogram Equalization (CLAHE)* to improve local contrast and reveal finer details. This enhancement allows for a more nuanced feature extraction through the same attention-based pipeline as the first branch, thereby achieving better feature representation while maintaining computational efficiency.
- The model combines features from both branches, leveraging the strengths of enhanced local contrast and advanced attention mechanisms to produce a comprehensive and robust feature set, which is subsequently fed into a classification module.
- In the extensive experiment with the PAD2020 and ISIC 2019 datasets, we assessed the proposed model and obtained an accuracy rate of 98.59%, of the PAD2020 surpassing the state-of-the-art performance by 2% and stable performance accuracy for the ISIC 2019 dataset. In addition, our model achieved stable performance accuracy for the ISIC-2019 dataset. This shows that the model is better at combining different attention mechanisms and feature improvement methods for accurate and useful skin cancer detection.

We organized the rest of the paper as follows: the literature review is included in Section 2, and datasets are described in Section 3. Methodology of the study is described in the Section 4, results are shown in the Section 5 and lastly, the conclusion is included in the Section 6.

2. Literature Review

Early detection is critical for effective treatment, as delayed diagnosis can lead to metastasis, increased morbidity, and higher mortality rates; however, traditional methods such as visual examination and biopsy are often subjective, time-consuming, and observer-dependent, prompting growing interest in machine learning—particularly deep learning—to enhance diagnostic accuracy and efficiency [1,3–5,7,8]. Convolutional Neural Networks (CNNs) have shown significant promise in detecting and classifying skin cancer from medical images [15,29,30]. Esteva et al. [46] utilized CNNs trained on over 129,000 clinical images to detect melanoma, achieving 95% sensitivity, 85% specificity, and 91% overall accuracy, comparable to dermatologists. Similarly, Tschandl et al. [47] achieved an AUC score of 0.94 in distinguishing benign from malignant lesions, and Brinker et al. [48] demonstrated enhanced accuracy using transfer learning. Although deep learning approaches have shown significant promise in detecting and classifying skin cancer, comparative studies are essential to evaluate the effectiveness of different CNN models designed for this purpose [15,17,49–51]. Several researchers have conducted comparisons to assess the efficacy of various transfer learning. Transfer learning, leveraging pre-trained CNNs for feature extraction, has further improved classification performance. Han et al. [52] achieved 89.1% accuracy in melanoma detection. Haenssle et al. [52] optimized a VGG-19 model to achieve an AUC of 0.86, while Codella et al. [53] refined an Inception-v3 model to reach an AUC of 0.93. These advancements underscore the transformative potential of deep learning in skin cancer detection. Codella et al. [53,54] compared Inception-V3, ResNet-50, and DenseNet-121, finding that DenseNet-121 achieved the highest classification accuracy with an AUC-ROC of 0.91. Tschandl et al. [53] evaluated five CNN models, including Inception-V3, DenseNet-121, and Xception, and discovered that Xception attained the best performance with an AUC-ROC of 0.93. Similarly, Kawahara et al. [55] used a pre-trained CNN to extract features and trained an SVM classifier, achieving an accuracy of 83.6%. Brinker et al. [48] tested five CNN models, reporting Inception-V3 as the top performer with an AUC of 0.90. Patel et al. [56] highlighted the potential of transfer learning, noting that the InceptionV3 model demonstrated high efficiency. Zhang et al. [57] found that DenseNet-121 achieved the highest AUC value of 0.95 on dermoscopic images. Zaidan et al. [58] emphasized the need for further studies using larger and diverse datasets to improve the robustness of CNN models for skin cancer detection. Recent research has explored alternative deep learning methods, such as Generative Adversarial Networks (GANs), for skin cancer detection and classification [59–61]. For instance, Bi et al. [62] utilized a GAN to generate artificial skin lesion images and employed a Convolutional Neural Network (CNN) to classify these images as benign or malignant, achieving a classification accuracy of 83%. Attention mechanisms have also shown promise in computer vision tasks, including image classification and object segmentation [44,63–66]. For example, reverse attention directs polyp segmentation by using characteristic masks as guiding signals [67], while co-attention modules enhance correlations between branches in Siamese networks for video frame analysis [68]. SENet employs global average pooling to model channel-wise feature dependencies through a squeeze-and-excitation process [63]. CBAM extends SENet by incorporating spatial information using larger kernels [64]. Coordinated Attention (CoorAttn) improves upon SENet by encoding long-range constraints and channel relationships across multiple dimensions [66]. Although these attention-based models enhance feature extraction, they require more trainable parameters and computational resources compared to SENet, making them resource-intensive for real-world applications. To enhance computational efficiency and address scaling requirements, natural language processing techniques like self-attention, particularly transformers, have been introduced to computer vision. The Vision Transformer (ViT) replaces traditional convolutional methods with a transformer encoder that leverages self-attention [65]. ViT processes input images by dividing them into patches and embedding them, resulting in large models with approximately 100 million parameters. We utilized the standard transformer to analyze sequences of image patches and capture inter-patch representations. However, the original transformer approach lacks inductive biases such as translation equivariance and locality inherent in CNNs, leading to suboptimal performance with limited training data [65]. To address this, we incorporated a ViT branch

into a CNN-based U-shaped architecture, enhancing long-range dependency modelling and contextual information extraction for skin lesion evaluation. While this approach improved feature representation, it also significantly increased the model's computational complexity and weight. Despite its effectiveness in reducing the number of learnable parameters, knowledge distillation involves complex parameter updates between teacher and student networks, resulting in high computational costs during training [69–71]. On the other hand, Atrous Spatial Pyramid Pooling (ASPP), a lightweight module, enables spatial sampling at multiple scales for better localization of segment borders. However, ASPP does not fully capture feature interactions. Depthwise Separable Convolutions (DWSCConv) offer an alternative to lightweight CNNs by replacing traditional convolutions with depthwise convolutions within channels and pointwise convolutions across channels, thereby significantly reducing parameter counts. Using DWSCConv, networks like MobileNetV2, MobileNetV3, EfficientNet, MnasNet, and ShuffleNet demonstrated decent performance [72–76]. This study also explores lightweight Vision Transformer (ViT) networks using sparse attention [41], random feature attention [77], and low-rank approximation [78] to reduce computational overhead. To address inductive bias limitations in ViTs, MobileViT combines convolution and transformer blocks into a hybrid architecture, offering improved parameter efficiency and inference speed [44]. Despite the use of EfficientNet and MobileNetV2 for skin lesion diagnosis [79,80], further research is needed to optimize parameter count and computational costs for mobile applications in skin lesion identification.

3. Dataset

This research utilizes two publicly accessible skin disease datasets, ISIC2019 [47,81] and PAD2020 [82], to develop and assess the proposed methodologies. Dermoscopy and cellphones are two conventional techniques for obtaining photos of skin lesions. As a result, the two datasets accurately represent contemporary imaging data for identifying skin lesions.

The ISIC2019 dataset comprises 25,331 dermoscopy photos categorized into eight classifications: actinic keratosis (ACK), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevus (NV), squamous cell carcinoma (SCC), and vascular lesion (VASC). Three (SCC, BCC, and ACK) are non-melanoma skin cancers. The most aggressive form of skin cancer is melanoma, resulting from the uncontrolled proliferation of pigment-producing cells. Vascular lesions (VASC) may be benign or malignant and necessitate vigilant observation over time. The other kinds of skin tumors are benign disorders. Figure 1 (a) shows the sample example of the ISIC2019 dataset. The PAD2020 dataset comprises 2,298 skin lesion photos categorized into six classes: BCC, MEL, BKL, ACK, NV, and SCC, obtained using cellphones. Figure 1 (b) shows the sample example of the PAD2020 dataset and can be downloaded for below link <https://www.kaggle.com/datasets/mahdavi1202/skin-cancer>. PAD2020 contains two fewer categories, DF and VASC, than ISIC2019 due to the absence of skin lesion photographs.

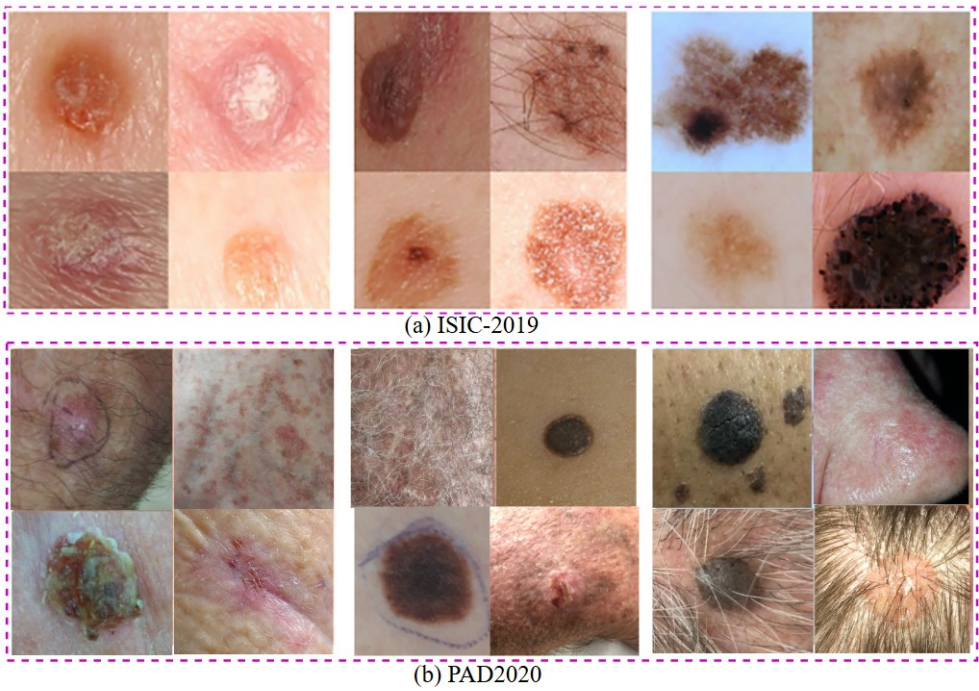


Figure 1. Example dataset samples: (a) ISIC2019 dataset [47,81], (b) PAD202 dataset [82].

4. Proposed Methodology

Figure 2 shows the proposed model architecture where we employed a hybrid attention-based two-stream deep neural network to develop a skin cancer recognition system which is the extension of the previous single stream model [45]. We began by preprocessing the input skin cancer dataset and employing augmentation techniques to effectively address overfitting issues. Our approach involves two distinct branches: the first extracts features from the original images, while the second branch extracts features from CLAHE-enhanced images. Each stream is designed with a convolutional layer integrated with a hybrid attention mechanism composed of three novel attention modules: Enhanced Separable Depthwise Convolution (SCAttn), stage attention, and branch attention-based deep neural networks. The primary objective of each stream is to capture both local and global hierarchical representations of the features, enabling a comprehensive understanding of the input data. The hybrid attention mechanism integrates three innovative attention modules—SCAttn [45], stage attention, and branch attention—to significantly enhance performance while maintaining a lightweight model architecture. The branch attention mechanism empowers the network to extract multi-level features from various layers, providing a deeper and more refined interpretation of the input data. The SCAttn module utilizes Depthwise Separable Convolutions (DWSCConv) to effectively incorporate global information, which substantially improves the model’s diagnostic accuracy. These attention mechanisms are remarkably efficient, consuming minimal or no learnable parameters, thereby ensuring the hybrid model remains compact without compromising its effectiveness. Furthermore, the stage attention mechanism employs a powerful self-attention strategy that integrates convolutional operations to retain inductive biases and reduce latency, inspired by the MobileViT architecture.

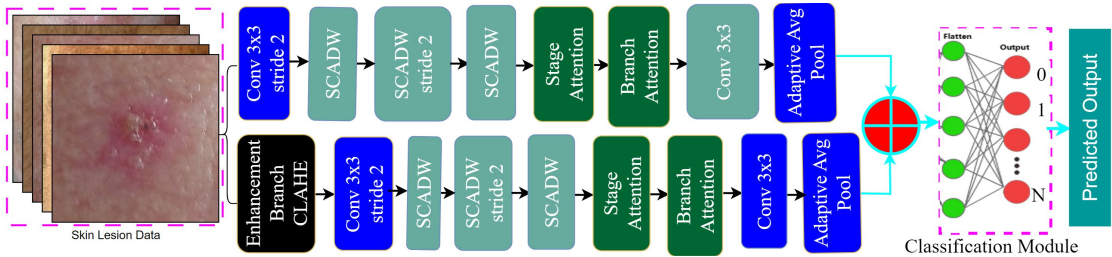


Figure 2. Workglow architecture of the proposed diagram

4.1. Image Pre-Processing

The ISIC2019 dataset contains numerous images with large dark regions, which can hinder the accurate evaluation of deep learning models. To address this, we adopt an adaptive cropping technique to focus on the relevant skin lesion areas. First, each image is converted to grayscale and then binarized using OpenCV's adaptive thresholding method with the mean adaptive approach (block size = 11, constant subtracted = 2) [83]. This method dynamically determines threshold values based on local neighborhood intensity. After binarization, we detect the curved surface to locate the central circular region. If the ratio of the circle's area to the entire image lies between 0.01 and 0.9, the circular region is cropped and preserved for further processing.

4.2. Data Balance

In this study, an unbalanced ratio is created to measure the imbalance problem. The unbalanced ratio represents a minor portion of the discrepancy between the number of pictures belonging to the ruling class and the minority class. For ISIC2019, the mismatch ratio is 53.9, and for PAD2020, it is 16.3. Previous studies have demonstrated how imbalanced ratios of this magnitude can significantly lower training performance [84]. Because the data isn't balanced, the small class might get a low validation result, but the overall validation result for all classes might be high. A useful way to deal with a very uneven collection is to balance the data by oversampling or undersampling each class. We simultaneously use both oversampling and undersampling to equalize the ISIC2019 and PAD2020 data. After sampling, we gathered 2500 pictures for the ISIC2019 dataset and 500 images for the PAD2020 dataset. This means that the total amount of data that has been adjusted is close to the total amount of unbalanced data, which is 20,000 for ISIC 2019 and 3,000 for PAD 2020. Oversampling to make new pictures from old ones uses random cuts, Gaussian blur, linear contrast, random translation, rotation, and shear on a small scale. Randomly selecting a set number of images can lead to a problem known as class overlapping. So, to fix this problem, we use instance hardness (IH), a flexible data analysis method [85].

4.3. Image Enhancement CLAHE

Contrast-Limited Adaptive Histogram Equalization (CLAHE) is an advanced image enhancement technique widely used in biomedical imaging to improve local contrast, particularly in regions with non-uniform illumination [86–90]. By enhancing local contrast, CLAHE helps reveal subtle structural details—such as lesion boundaries—critical for medical image analysis. Unlike global histogram equalization, CLAHE operates by dividing the image into small non-overlapping regions, or tiles, typically using an 8×8 grid [91,92]. Within each tile, the local histogram is computed and clipped to a predefined clip limit before being redistributed. This clip limit, defined as $CL = N_{cl} \times N_{avg}$, controls contrast enhancement by preventing over-amplification of noise. Here, N_{cl} is the normalized clip limit (commonly set to 2), and $N_{avg} = \frac{N_x \times N_y}{N_{gray}}$ represents the average number of pixels per gray level, where N_x and N_y are the tile's dimensions and N_{gray} is the number of gray levels. If the frequency n_j of a gray level j exceeds the clip limit CL , the excess pixels are clipped and redistributed uniformly across all gray levels to preserve the overall intensity distribution. This generates a contrast-limited histogram for each tile. To mitigate block artifacts that may occur due to independent tile processing, CLAHE uses bilinear interpolation to seamlessly blend the tiles, resulting in a more natural and continuous image enhancement [88,89]. By enhancing local contrast without amplifying noise excessively, CLAHE has proven especially effective when integrated into deep learning pipelines for medical imaging tasks, such as lesion detection and skin disease classification.

4.4. CTH Module

The Convolution-Transformer Hybrid (CTH) module, as shown in Figure 3 (a), integrates sequential components designed to process feature maps before and after encoding. The CTH block is composed of a multi-head self-attention (MHSA) component followed by a feed-forward neural network (FFN) layer, commonly referred to as a vision transformer (ViT) [45]. This ViT component

effectively captures long-range spatial relationships within the feature maps. To address the need for capturing local features, the CTH block integrates standard convolution operations before and after the ViT module. This addition ensures that both local and global features are captured, as illustrated in Figure 3 (a). Furthermore, the CTH block includes a skip connection that connects the input and output, allowing for the preservation of low-level information from earlier layers. The Stage Attention module operates alongside the CTH block to extensively reorganize the feature maps through downsampling. The CTH block extracts the modified features and uses them for further processing. This integration of attention mechanisms and convolution allows the model to efficiently learn hierarchical features. The following steps outline the operations performed within the CTH block.

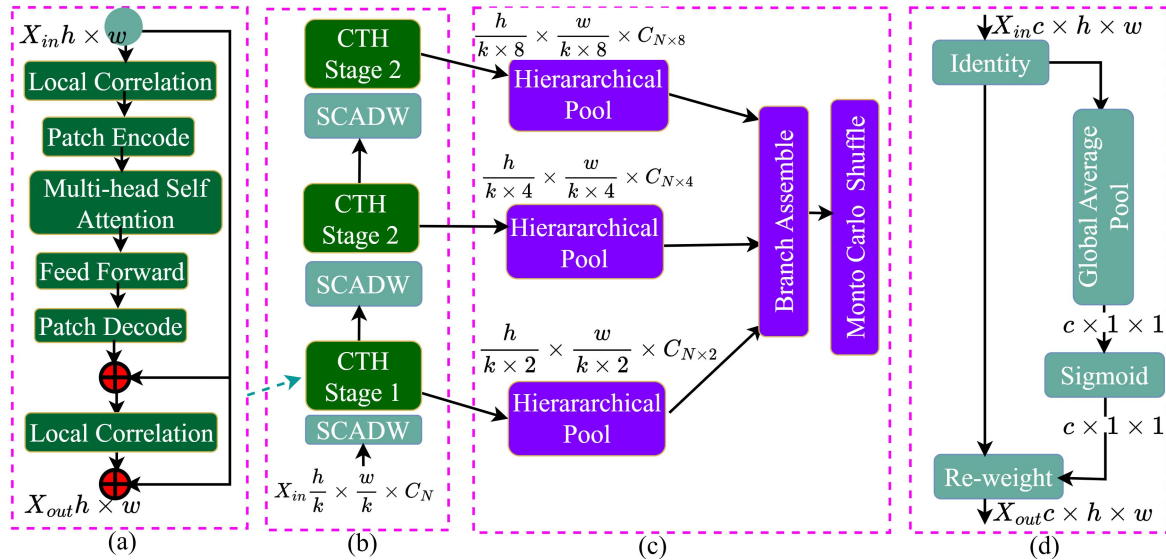


Figure 3. Details of (a) CTH Module (b) Stage attention module (c) Batch attention module and (d) SCADW module [45]

Step 1: Local Correlation (Convolutional Feature Processing) Before feeding the input into the ViT module, local correlations are applied to the input feature map. This is achieved using convolutional layers that preserve local feature information.

$$X_{\text{local}} = \text{LocalCorr}(X_{\text{in}}) + X_{\text{in}} \quad (1)$$

Where, X_{in} is the input feature map. X_{local} represents the output after local correlation is applied.

Step 2: Multi-Head Self-Attention (MHSA)

The next step applies the MHSA module to the input feature map to capture long-range dependencies between the patches:

$$Y = \text{MHSA}(\text{Encode}(X_{\text{local}})) \quad (2)$$

Where, $\text{Encode}(X_{\text{local}})$ encodes the input feature map into patches for attention computation. Y is the feature map after applying the MHSA module.

Step 3: Feed-Forward Network (FFN)

The output from the MHSA is passed through a Feed-Forward Network (FFN), followed by a layer normalization to further refine the features:

$$Z = \text{Decode}(\text{LayerNorm}(\text{FFN}(Y))) \quad (3)$$

Where LayerNorm normalizes the features before applying the FFN. Z is the output after decoding and normalization.

Step 4: Local Correlation and Skip Connection

Finally, the output of the ViT module, combined with the original input, is processed through another convolutional layer for final feature extraction:

$$X_{out} = \text{LocalCorr}(Z + X_{in}) + X_{in} \quad (4)$$

Where X_{out} is the final output feature map. This step ensures that local features are captured while retaining the global contextual information learned from the ViT. This hybrid approach allows the CTH module to effectively capture both local and global feature representations, leading to a more robust model that can handle a wide range of tasks.

4.5. SCADW

The block modifies the DWSCConv with the SCAttn technique, resulting in a new SCADW block structure shown in Figure 3 (d). The SCADW block is structured to enhance the performance of the HierAttn architecture while maintaining efficiency [45]. The overall structure of the SCADW block can be summarized as follows: This structure allows the SCADW block to effectively learn and represent features that are crucial for skin lesion classification while ensuring that the model remains efficient and suitable for deployment in resource-constrained environments.

The benefits of the SCADW model are that it enhances feature representation by combining depthwise separable convolution with the SCAttn module, capturing local and global features effectively. It maintains a compact model size, making it suitable for mobile devices with limited computational resources. Here are the details regarding the structure of SCADW.

1. **Depthwise Separable Convolution(DWSCConv):** The SCADW block begins with a depthwise separable convolution which consists of two main operations that are- 1) Depthwise Convolution operation that applies a single filter to each input channel separately, which helps in reducing the number of parameters and computational cost compared to standard convolution, 2) Pointwise Convolution that follows the depthwise convolution. Pointwise convolution (1x1 convolution) is applied to combine the outputs from the depthwise convolution across all channels. DWSCConv reduces computational complexity by separating convolution into depthwise and pointwise operations, maintaining performance in tasks like image classification and object detection, and allowing for more complex architecture design without increasing computational requirements.
2. **Integration of SCAttn Module:** After the depthwise separable convolution, the SCADW block integrates the SCAttn (Same Channel Attention) module. The SCAttn module utilizes global average pooling to extract global features from the output of the depthwise convolution. This process allows the model to focus on the most relevant features without redundantly processing channel-wise information. SCAttn uses global average pooling to extract global features from the depthwise convolution output, improving feature representation. It reduces redundant operations, lowers computational costs, and is compatible with lightweight models, making it ideal for mobile and edge device applications.
3. **Output:** The output of the SCADW block is a feature map that has been enhanced through the attention mechanism, allowing the model to capture important global context while preserving local feature information [45].

4.6. Stage Attention

Stage attention is key to the HierAttn architecture for skin lesion diagnosis which shows in the 3 (b). It combines a SCADW block with a convolution transformer hybrid (CTH) block to learn local and global features effectively [45]. This module downsamples feature maps, aggregates local features, and applies attention at different stages to enhance focus on relevant features. Stage attention, inspired by the MobileViT architecture, ensures efficient hierarchical learning with low latency, which is critical for mobile applications. Overall, stage attention improves the HierAttn network's classification accuracy and AUC scores by enabling multi-level feature representation.

4.7. Branch Attention

To enhance feature interactions during training, Zhang et al. calculated three additional losses inside intermediate phases[93]. We used *branch attention*, a deep supervision method that uses hierarchical pooling to make tensors from later stages smaller [45]. It can be seen in the lower 3 (c). This approach uses tensor assembly to enhance the interactions of the shrunk features and learn the feature hierarchy from the built tensors. Furthermore, hierarchical pooling optimizes tensor representations by preserving the varying sizes of pooling results. This allows for the generation of a large tensor size ($C \times 5 \times 5$) while capturing global representations, resulting in a lower tensor size ($C \times 1 \times 1$). A medium output tensor of dimensions ($C \times 3 \times 3$) is included as a buffer layer to maintain the integrity of both local and global characteristics. Following the channel-wise assembly of the pooled tensors from various branches, the ensembled tensors are next combined pixel-wise. Once the branch attention is complete, the Monte Carlo technique reorganizes the tensors pixel by pixel. Each stage attention block, also known as the key learning stage, implements the branch attention operation as a specific learning phase. Significantly, this pragmatic design avoids introducing any extra loss during the intermediate stage, thereby requiring relatively lower computational resources than the traditional deep supervision approach.

4.8. Stream-1: Original Image Based Feature

We implemented an initial convolutional layer to handle the preprocessed image in the first branch. Then, three SCAttn (Enhanced Separable Depthwise Convolution) modules with 1, 2, and 1 stride sizes were sequentially integrated [45]. This setup improves the model's capacity to grasp both detailed and wide contextual information by facilitating efficient feature extraction at different receptive fields. The computational efficiency is maintained by each SCAttn module, which efficiently concentrates on key spatial features, thanks to its depthwise separable structure. The output from the SCAttn modules is then processed through the stage attention mechanism, which further refines the feature maps by dynamically adjusting the focus on important regions across different stages of the network. This stage attention ensures that both local and global feature representations are optimally balanced, reducing redundancy and enhancing feature diversity. After the features are improved, they are sent to the branch attention module, which groups and puts together features from different branches in a hierarchical way. This step greatly improves how features interact with each other and lets the model learn complicated hierarchical representations, bringing together specific local patterns with important global contexts. The hybrid attention modules process the combined features and then input them into an additional convolutional layer to capture and refine abstract concepts. Lastly, the final feature representation, $F_{GStream}$ is created by applying an average pooling layer. The model's diagnostic accuracy and resilience are enhanced by the first branch's ability to extract a rich set of features, which captures intricate local details while simultaneously preserving a knowledge of the overall spatial structure. The thorough integration of numerous attention mechanisms achieves this.

4.9. Stream-2: Enhanced Image-Based Feature

We improved the input image using the CLAHE method in the second branch. By altering the histogram of each small region, CLAHE successfully enhances local contrast and displays finer details in the skin lesion photos. To give the model a more complete input, this improvement is essential for bringing attention to delicate patterns and features that may go unnoticed during regular preprocessing. To make more accurate feature representations, CLAHE boosts the effectiveness of each layer of feature extraction by focusing on key details while keeping edge information. Extracting spatial characteristics from the CLAHE-enhanced image begins with its passage through a convolutional layer. Three SCAttn (Enhanced Separable Depthwise Convolution) modules, each having a stride size of 1, 2, and 1, are then fed this output in that order. Using the depthwise separable convolutions to efficiently zero in on important spatial details with minimal computational overhead, the integration of CLAHE enhancement with SCAttn modules takes advantage of the enhanced local contrast to

extract more effective features. The stage attention mechanism processes the output from the SCAttn modules and dynamically shifts the emphasis to crucial regions throughout stages, ensuring a balance between refined local details and broader contextual information. Because it enables the model to focus on the most informative portions of the image, this step is especially beneficial when coupled with the CLAHE-enhanced input. The next step is to use the branch attention module to improve the features. This module hierarchically assembles and pools features from multiple levels. By studying a complicated hierarchy that captures both important global contexts and specific local patterns, this module improves the interactions between features even further. The output is refined and abstracted further by passing it through an additional convolutional layer after integrating these features using the hybrid attention techniques. The last step in creating the feature representation, $F_{CStream}$, is to apply an average pooling layer. With the help of CLAHE improvement and several attention modules, this branch can gather improved local textures as well as global structural patterns, leading to a more accurate and resilient model.

4.10. Feature Concatenation and Classification

To further enhance the model's feature representation, we combined the two streams' features by concatenating them, as shown in Equation 5:

$$F_{Concat} = F_{GStream} \oplus F_{CStream} \quad (5)$$

where F_{Concat} represents the concatenated feature vector fed into the classification module. This concatenation of features from the two branches, $F_{GStream}$ and $F_{CStream}$, effectively integrates complementary information from both the original and CLAHE-enhanced images. The first branch, $F_{GStream}$, captures the fundamental features from the preprocessed images, providing a solid base of local and global hierarchical representations. Meanwhile, the second branch, $F_{CStream}$, contributes enhanced features that emphasize critical local contrast and subtle patterns revealed by CLAHE. By merging these two streams, the model benefits from a richer, more comprehensive feature set that includes both the original spatial structures and the enhanced details. This fusion enables the model to leverage the strengths of both inputs, improving its ability to differentiate between skin lesions with subtle variations. The combined feature representation, F_{Concat} , offers a more robust and discriminative input for the classification module, ultimately enhancing the overall diagnostic accuracy and effectiveness of the proposed model.

4.11. Classification and Training Procedure

After feature extraction through the dual-branch network, we employed a robust classification module to produce a probabilistic map for skin cancer detection. This module takes the concatenated feature vector, F_{Concat} , and processes it through a series of fully connected layers. Each layer refines the features by learning complex, non-linear relationships, which are critical for distinguishing between different skin lesion types. The output layer of the classification module uses a softmax activation function to generate a probabilistic map, providing the likelihood of each input belonging to different skin cancer categories. This probabilistic map not only facilitates accurate disease detection but also offers insights into the model's confidence in its predictions.

The training procedure is designed to optimize the classification performance by minimizing a cross-entropy loss function, which measures the divergence between the predicted probabilities and the true class labels. We employed a stochastic gradient descent (SGD) optimizer with momentum to accelerate convergence and avoid local minima. The training process includes data augmentation techniques and regularization methods, such as dropout, to prevent overfitting and improve the model's generalization capability. The model is trained iteratively, continuously updating weights based on backpropagation until it reaches optimal performance, thus ensuring reliable and precise skin cancer detection.

5. Experimental Result

We evaluated the proposed model using various datasets with various modalities. Based on the old procedure, we divided the dataset into two parts: training and testing. Additionally, we display our environmental settings and the effectiveness of our information processing below.

5.0.1. Environmental Setup and Evaluation Metrics

The system was constructed using a GeForce RTX 4090 24GB graphics processing unit (GPU), CUDA 10.7, NVIDIA driver 515, and 64GB of random access memory (RAM). The training experiment had a learning rate of 0.005 and a batch size of 32. It was executed for 300 epochs utilizing the Adam optimizer on the RTX 4090 architecture. To achieve optimized graph convolution and attention with minimal computational expense, the Python environment used OpenCV, Pickle, Pandas, and PyTorch for model construction[94]. These packages, in addition to a few others[95,96], enabled the initial processing of data and the building of models. The assessment measure for assessing the accuracy of the proposed identification model, analogous to Equation 6, is calculated using TN for true negatives, TP for true positives, FN for false negatives, and FP for false positives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

We assessed the proposed model using two datasets and computed the accuracy of the performance for many models for each dataset.

5.1. Ablation Study

The ablation study evaluates the impact of applying augmentation before and after splitting the dataset on model performance shows in Table 1. In the augmentation before splitting approach, augmentation is applied to the entire dataset before the train-test split, ensuring that both the training and testing sets benefit from the enhanced features. This configuration showed the highest performance, achieving 98.59% accuracy on the PAD2020 dataset and 83.00% (average) accuracy on the ISIC-2019 dataset, with a peak of 88.62% in the best fold. This method leverages the full potential of data augmentation, improving the feature representation for both training and testing phases. In contrast, the augmentation after splitting, where augmentation is applied only to the training set post-split, resulted in lower performance. The accuracy dropped to 70.00% on PAD2020 and 62.65% on ISIC-2019, indicating that applying augmentation only to the training data after the split fails to fully exploit the benefits of enhanced feature extraction across the entire dataset. This highlights the importance of applying augmentation before the data split to achieve superior performance and feature representation.

Table 1. Performance Comparison of Ablation Study with Augmentation Models on PAD2020 and ISIC-2019 Datasets.

Ablation Study	Augmentation Models	Mod-	Stream-1	Stream-2	PAD2020 (Accuracy %)	ISIC-2019 (Accuracy %)
Single Stream	First Augmentation → Then Splitting train-test		Stream-1	-	94.22	82.00
Dual Stream	First Splitting train-test → Then Augmentation only on train set		Stream-1	Stream-2	70.00	62.65
Proposed Dual Stream	First Augmentation → Then Splitting train-test		Stream-1	Stream-2	98.59	83.00 (Average), 88.62 (Best Fold)

5.2. Performance Accuracy for the PAD2020 Dataset

Table 2 shows the performance metrics of the proposed model across ten validation folds using the PAD2020 dataset. The evaluation includes Best Accuracy, Recall, Specificity, Precision, and F1 Score for each fold. The model demonstrates consistently high accuracy (average 98.59%) and specificity (average 99.71%), indicating a strong ability to correctly identify true negatives. The average recall (79.44%), precision (80.18%), and F1 score (79.78%) reflect a balanced performance in detecting positive cases, confirming the model’s effectiveness and robustness in Parkinson’s disease classification.

Table 2. Performance Metrics for each validation fold for PAD2020 dataset.

K-Fold	BestAccuracy	BestRecall	BestSpecificity	BestPrecision	BestF1Score
val_1	98.31	78.68	99.66	79.13	78.90
val_2	98.55	79.57	99.71	79.44	79.50
val_3	98.54	79.20	99.70	80.19	79.67
val_4	98.87	79.64	99.76	81.40	80.44
val_5	98.68	79.68	99.73	80.18	79.92
val_6	98.49	79.33	99.69	79.99	79.65
val_7	98.27	78.81	99.64	79.33	79.06
val_8	99.02	80.69	99.78	81.85	81.24
val_9	98.76	80.48	99.75	80.24	80.35
val_10	98.38	78.30	99.66	80.09	79.11
Average	98.59	79.44	99.71	80.18	79.78

Table 3 provides a comparative analysis of recent studies focusing on various skin cancer datasets, specifically highlighting different models and their performance in terms of accuracy. The datasets considered include PAD2020, which shows that VGG16 models achieved an accuracy of 70.83%. Among the models applied exclusively to the PAD2020 dataset, the MobileNetV2 and MobileNetV3L models achieved high accuracy rates of 91.44% and 91.78%, respectively. ShuffleNetV2_1.x and EfficientNet_b0 also performed well, with accuracies around 91.89% and 91.13%. Meanwhile, more advanced methodologies, such as the HierAttn series models, reported the highest accuracy, with HierAttn_s reaching 91.22% [45], demonstrating attention mechanisms’ potential in enhancing model performance. This comprehensive comparison underscores the efficacy of various deep-learning models and techniques in enhancing skin cancer detection accuracy. Models like ResNet101, MnasNet1.0, and RegNetY6.4gf show an appropriate combination of computational efficiency and high accuracy, rendering them viable options for clinical applications and automated diagnostics in dermatology.

Table 3. State of the art comparison of the proposed model for the PAD202 datasets.

Reference	Published Year	Methodology	Dataset	Performance Accuracy [%]
VGG16		VGG16	PAD2020	70.83%
ResNet101		ResNet101	PAD2020	90.67%
MobileNetV2		MobileNetV2	PAD2020	91.44%
MobileViT_s		MobileViT_s	PAD2020	89.22%
MobileNetV3L		MobileNetV3L	PAD2020	91.78%
ShuffleNetV2_1.x		ShuffleNetV2_1.x	PAD2020	91.89%
RegNetY6.4gf		RegNetY6.4gf	PAD2020	90.67%
MnasNet1.0		MnasNet1.0	PAD2020	91.33%
EfficientNet_b0		EfficientNet_b0	PAD2020	91.13%
Ensembled Model		Ensembled Model	PAD2020	94.11%
HierAttn_xs [45]	2024	HierAttn_xs	PAD2020	90.11%
HierAttn_s [45]	2024	HierAttn_s	PAD2020	91.22%
Proposed Method	-	Two Stream Model	PAD2020	98.59%

5.3. Performance Accuracy for the ISIC-2019 Dataset

Table 4 demonstrated the fold-wise performance accuracy for the ISIC-2019 dataset. To prove the stability and generalizability of the proposed model, we also evaluated the ISIC-2019 datasets. We achieved 82.00% accuracy for the single stream in our own experiment, which is a similar model to the existing article [45], and 83.00% for the proposed dual stream. This performance proves that our proposed model also produces stable performance accuracy with other datasets.

Table 4. Performance Metrics for Each Validation Fold for the ISIC2019 dataset

K-Fold	BestAccuracy	BestRecall	BestSpecificity	BestPrecision	BestF1Score
val_1	83.44	83.29	97.64	83.49	83.34
val_2	87.65	87.54	98.24	87.65	87.49
val_3	79.85	79.64	97.12	79.80	79.49
val_4	81.84	81.71	97.41	82.14	81.82
val_5	84.14	84.01	97.73	84.48	84.08
val_6	77.49	77.23	96.79	77.78	77.35
val_7	82.49	82.37	97.50	82.31	82.22
val_8	88.63	88.55	98.37	88.94	88.60
val_9	86.25	86.02	98.04	86.45	86.14
val_10	78.69	77.55	96.97	78.01	77.66
Average	83.05	82.79	97.58	83.11	82.82

Table 5 compares several state-of-the-art models applied to various skin lesion datasets, with a focus on the ISIC 2019 dataset. Models such as CNN, ResNet50, Xception, DenseNet201, MobileNetV2, and others are evaluated based on different training configurations, including splitting first, followed by augmentation, or vice versa. Performance metrics include accuracy, F1-score, and specificity. The proposed dual-stream model outperforms others with 88.62% accuracy and 98.37% specificity, demonstrating superior classification efficiency and minimizing false positives. In comparison, our experiments with a single-stream model achieved 82.00% accuracy, similar to the architecture used by Dai et al. [45], which reported 96.70% accuracy. This discrepancy suggests potential differences in experimental conditions or dataset variations. However, the dual-stream model showed a slight improvement. This comparison highlights the proposed model’s stability cross different datasets, emphasising the importance of model architecture choices in achieving higher accuracy and specificity in skin lesion classification.

Table 5. State-of-the-art comparison with ISIC-2019 dataset

Authors and Paper	Dataset	Model	Published Year	Performance Accuracy with for first splitting then Augmentation only training set	Performance Accuracy with for first augmentation then splitting their training and testing set	Specificity
Inthiyaz et al. [97]	Xiangya-Derm	CNN	2023	-	AUC = 0.87	-
Alwakid et al. [98]	HAM10000	CNN, ResNet50	2023	-	F1-score = 0.859 (CNN), 0.852 (ResNet50)	-
Aljohani and Turki [99]	ISIC 2019	Xception, DenseNet201, ResNet50V2, MobileNetV2, 2022 VGG16, VGG19, and GoogleNet CNN,	2022	-	Accuracy = 76.09%	-
Bechelli and Delhom-melle [100]	HAM10000	Xcep-tion, ResNet50	2022	-	Accuracy = 88% (VGG16)	-
Demir et al. [101]	ISIC archive	ResNet101 and Incep-tionV3	2019	-	F1-score = 84.09% (ResNet101), 87.42% (Incep-tionV3)	-
Kaggle Compt. [102]	ISIC2018	Top 10 model Average	2020	-	Accuracy = 86.7%	-
Gouda et al. [103]	ISIC 2018	CNN	2022	62.50	Accuracy = 83.2%	-
Proposed Model	ISIC 2018	Duel Strem	-	62.65(Best Fold)	Accuracy = 88.62% (Best Fold)	98.37 (Best Fold)

5.4. Dicussion

The proposed Hybrid Attention Dual-Stream Deep Learning Model introduces a lightweight and effective architecture for skin cancer detection. It leverages a dual-branch design where the first branch processes the raw image through multi-stage attention modules (SCAttn, stage attention, and branch attention), while the second branch uses CLAHE-enhanced images to extract fine-grained local details. Both branches’ features are fused and passed through a unified classification module under a single loss function, ensuring end-to-end learning and robust generalization. To evaluate the model’s effectiveness, we conducted experiments on two publicly available datasets: PAD2020 and ISIC-2019. The proposed model achieved a state-of-the-art performance of 98.59% accuracy on PAD2020, significantly surpassing prior models such as MobileNetV3L (91.78%), EfficientNet_b0

(91.13%), and even ensembled models (94.11%). The ablation study further revealed that applying augmentation before the dataset split considerably improves accuracy, supporting the model's ability to extract rich, representative features across training and testing phases. On the ISIC-2019 dataset, the proposed model reached an average accuracy of 83.00% with a best fold performance of 88.62% and specificity of 98.37%, outperforming several existing deep CNN models including Xception, DenseNet201, and ResNet50V2, which reported lower accuracies ranging from 76.09% to 86.7% in previous studies. Although Dai et al. reported 96.70% with a different experimental setting, our dual-stream architecture demonstrated more stable cross-validation performance, confirming the model's generalizability.

In addition to its medical imaging applications, the proposed hybrid attention dual-stream model has strong implications for Future Internet technologies, particularly in the domains of edge AI and smart healthcare systems. The lightweight and modular design of the model makes it suitable for deployment on resource-constrained devices such as mobile phones and embedded systems, enabling real-time analysis and feedback without relying on cloud-based computation. This supports intelligent human-machine interfaces, where patients or clinicians can interact with AI-powered tools directly through connected devices. Moreover, integrating this model within IoT-based medical networks could facilitate remote diagnostics and teledermatology, especially in underserved or rural areas. In such scenarios, the model enables efficient data processing at the edge, reducing latency and bandwidth usage—both of which are critical factors in the design of Future Internet architectures for healthcare. These capabilities align well with the vision of the Future Internet, where distributed intelligence, networked sensors, and user-centric interfaces drive personalized and accessible medical care.

6. Conclusions

This study tackles the significant difficulty of accurate and prompt skin cancer diagnosis by introducing an innovative hybrid attention dual-stream deep learning model for the identification of skin lesions. Our model overcomes the limitations of existing ensemble convolutional neural networks (CNNs), which are often too large and inefficient for contextual processing, as well as lightweight networks like MobileNetV3 and EfficientNet, which have inadequate feature representation depth. Our approach integrates two complementary branches: one that extracts local and global features using Enhanced Separable Depthwise Convolution (SCAttn), stage attention, and branch attention modules, and another that enhances images using CLAHE to capture more effective features. This dual-stream model, which combines multiple attention mechanisms and feature enhancement techniques, significantly improves the robustness and accuracy of skin lesion detection. Clinically, this advancement enhances early detection and diagnostic accuracy, which are critical factors in improving patient outcomes. The model's ability to provide a comprehensive, hierarchical feature set makes it a promising tool for dermatologists, offering more reliable and efficient skin cancer diagnosis. Moving forward, future work will focus on refining the model's architecture for handling more complex datasets, exploring additional attention mechanisms, and expanding its applicability to other medical imaging tasks, further improving its clinical utility.

Author Contributions: Conceptualization, A.S.M.M.,K.H.; Methodology, A.S.M.M.,K.H.; Software, A.S.M.M.,K.H.; Validation,A.S.M.M.,K.H.; Formal analysis, A.S.M.M., K.H., and J.S.; Investigation, A.S.M.M., K.H., and J.S.; Resources, A.S.M.M.,K.H., J.S.; Writing—original draft preparation, A.S.M.M.; Writing—review and editing, K.H. and J.S.; Visualization, K.H.; Supervision, J.S.; Project administration, A.S.M.M. and J.S.; Funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: To be added

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

Abbreviations

The following abbreviations are used in this manuscript:

AlexNet	Alex Krizhevsky’s Network
AUC	Area Under the Curve
CNN	Convolutional Neural Networks
DenseNet121	Densely Connected Network 121
DenseNet201	Densely Connected Network 201
GAN	Generative Adversarial Network
InceptionResNetV2	Inception + ResNet Variant 2
InceptionV3	Inception Variant 3
MobileNetV2	Mobile Network Variant 2
ReLU	Rectified Linear Unit
ResNet50	Residual Network 50
ResNet50V2	Residual Network 50 Variant 2
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
VGG16	Visual Geometry Group 16
Xception	Extreme Inception

References

1. Nawaz, K.; Zaniab, A.; Shabir, I.; Li, J.; Wang, Y.; Mahmood, T.; Rehman, A. Skin cancer detection using dermoscopic images with convolutional neural network. *Scientific Reports* **2025**, *15*, 7252.
2. US Department of Health and Human Services. The Surgeon General’s Call to Action to Prevent Skin Cancer. Washington (DC): Office of the Surgeon General (US), 2014. Skin Cancer as a Major Public Health Problem, Accessed: 2023-08-22, Available from: <https://www.ncbi.nlm.nih.gov/books/NBK247164/>.
3. Mazhar, F.; Aslam, N.; Naeem, A.; Ahmad, H.; Fuzail, M.; Imran, M. Enhanced Diagnosis of Skin Cancer from Dermoscopic Images Using Alignment Optimized Convolutional Neural Networks and Grey Wolf Optimization. *Journal of Computing Theories and Applications* **2025**, *2*.
4. Farooq, A.B.; Akbar, S.; ul Ain, Q.; Naqvi, Z.; Urooj, F. Skin Cancer Detection and Classification Using Explainable Artificial Intelligence for Unbalanced Data: State of the Art. *Explainable Artificial Intelligence in Medical Imaging* **2025**, pp. 124–146.
5. Khullar, V.; Kaur, P.; Gargish, S.; Mishra, A.M.; Singh, P.; Diwakar, M.; Bijalwan, A.; Gupta, I. Minimal sourced and lightweight federated transfer learning models for skin cancer detection. *Scientific Reports* **2025**, *15*, 2605.
6. Nguyen, A.T.P.; Jewel, R.M.; Akter, A. Comparative Analysis of Machine Learning Models for Automated Skin Cancer Detection: Advancements in Diagnostic Accuracy and AI Integration. *The American Journal of Medical Sciences and Pharmaceutical Research* **2025**, *7*, 15–26.
7. Dorathi Jayaseeli, J.; Briskilal, J.; Fancy, C.; Vaitheeshwaran, V.; Patibandla, R.L.; Syed, K.; Swain, A.K. An intelligent framework for skin cancer detection and classification using fusion of Squeeze-Excitation-DenseNet with Metaheuristic-driven ensemble deep learning models. *Scientific Reports* **2025**, *15*, 7425.
8. Akter, M.; Khatun, R.; Talukder, M.A.; Islam, M.M.; Uddin, M.A.; Ahamed, M.K.U.; Khraisat, A. An Integrated Deep Learning Model for Skin Cancer Detection Using Hybrid Feature Fusion Technique. *Biomedical Materials & Devices* **2025**, pp. 1–15.
9. Kaur, R.; GholamHosseini, H.; Lindén, M. Advanced Deep Learning Models for Melanoma Diagnosis in Computer-Aided Skin Cancer Detection. *Sensors* **2025**, *25*, 594.
10. Sobahi, N.; Alhawsawi, A.M.; Damoom, M.M.; Sengur, A. Extreme Learning Machine-Mixer: An Alternative to Multilayer Perceptron-Mixer and Its Application in Skin Cancer Detection Based on Dermoscopy Images. *Arabian Journal for Science and Engineering* **2025**, pp. 1–16.
11. Skin Cancer Foundation. Skin Cancer Facts & Statistics. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>, n.d. Accessed: 2025-06-13.
12. Zhang, B.; Zhou, X.; Luo, Y.; Zhang, H.; Yang, H.; Ma, J.; Ma, L. Opportunities and challenges: Classification of skin disease based on deep learning. *Chinese Journal of Mechanical Engineering* **2021**, *34*, 1–14.
13. de Zarzà, I.; de Curtò, J.; Hernández-Orallo, E.; Calafate, C.T. Cascading and Ensemble Techniques in Deep Learning. *Electronics* **2023**, *12*. <https://doi.org/10.3390/electronics12153354>.

14. Loescher, L.J.; Janda, M.; Soyer, H.P.; Shea, K.; Curiel-Lewandrowski, C. Advances in skin cancer early detection and diagnosis. In Proceedings of the Seminars in oncology nursing. Elsevier, 2013, Vol. 29, pp. 170–181.
15. Naqvi, M.; Gilani, S.Q.; Syed, T.; Marques, O.; Kim, H.C. Skin Cancer Detection Using Deep Learning — A Review. *Diagnostics* **2023**, *13*. <https://doi.org/10.3390/diagnostics13111911>.
16. Sethanan, K.; Pitakaso, R.; Srichok, T.; Khonjun, S.; Thannipat, P.; Wanram, S.; Boonmee, C.; Gonwirat, S.; Enkvetchakul, P.; Kaewta, C.; et al. Double AMIS-ensemble deep learning for skin cancer classification. *Expert Systems with Applications* **2023**, *234*, 121047. <https://doi.org/10.1016/j.eswa.2023.121047>.
17. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data* **2021**, *8*, 1–74.
18. Miah, A.S.M.; Mamunur Rashid, M.; Redwanur Rahman, M.; Tofayel Hossain, M.; Shahidujjaman Sujon, M.; Nawal, N.; Hasan, M.; Shin, J. Alzheimer's disease detection using CNN based on effective dimensionality reduction approach. In Proceedings of the Intelligent Computing and Optimization: Proceedings of the 3rd International Conference on Intelligent Computing and Optimization 2020 (ICO 2020). Springer, 2021, pp. 801–811.
19. Hassan, N.; Musa Miah, A.S.; Shin, J. Residual-Based Multi-Stage Deep Learning Framework for Computer-Aided Alzheimer's Disease Detection. *Journal of Imaging* **2024**, *10*. <https://doi.org/10.3390/jimaging10060141>.
20. Shin, J.; Miah, A.S.M.; Hirooka, K.; Hasan, M.A.M.; Maniruzzaman, M. Parkinson Disease Detection Based on In-air Dynamics Feature Extraction and Selection Using Machine Learning. *arXiv preprint arXiv:2412.17849* **2024**.
21. Avilés-Izquierdo, J.A.; Molina-López, I.; Rodríguez-Lomba, E.; Marquez-Rodas, I.; Suarez-Fernandez, R.; Lazaro-Ochaita, P. Who detects melanoma? Impact of detection patterns on characteristics and prognosis of patients with melanoma. *Journal of the American Academy of Dermatology* **2016**, *75*, 967–974.
22. Sonkin, D.; Thomas, A.; Teicher, B.A. Cancer treatments: Past, present, and future. *Cancer Genetics* **2024**, *286–287*, 18–24. <https://doi.org/10.1016/j.cancergen.2024.06.002>.
23. Liu, H.; Dilger, J.P. Different strategies for cancer treatment: Targeting cancer cells or their neighbors? *Chinese Journal of Cancer Research* **2025**, *37*, 289.
24. Joshi, R.M.; Telang, B.; Soni, G.; Khalife, A. Overview of perspectives on cancer, newer therapies, and future directions. *Oncology and Translational Medicine* **2024**, *10*, 105–109.
25. Celebi, M.E.; Iyatomi, H.; Schaefer, G.; Stoecker, W.V. Lesion border detection in dermoscopy images. *Computerized medical imaging and graphics* **2009**, *33*, 148–153.
26. Peruch, F.; Bogó, F.; Bonazza, M.; Cappelleri, V.M.; Peserico, E. Simpler, faster, more accurate melanocytic lesion segmentation through meds. *IEEE Transactions on Biomedical Engineering* **2013**, *61*, 557–565.
27. Yamashita, R.; Nishio, M.; Do, R.K.G.; et al. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights into Imaging* **2018**, *9*, 611–629. <https://doi.org/10.1007/s13244-018-0639-9>.
28. Taye, M.M. Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation* **2023**, *11*, 52. <https://doi.org/10.3390/computation11030052>.
29. Raval, D.; Undavia, J.N. A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images. *Healthcare Analytics* **2023**, *3*, 100199. <https://doi.org/10.1016/j.health.2023.100199>.
30. Iqbal, S.; Qureshi, A.N.; Li, J.; et al. On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks. *Archive of Computational Methods in Engineering* **2023**, *30*, 3173–3233. <https://doi.org/10.1007/s11831-023-09899-9>.
31. Miah, A.S.M.; Hasan, M.A.M.; Shin, J. Dynamic Hand Gesture Recognition using Multi-Branch Attention Based Graph and General Deep Learning Model. *IEEE Access* **2023**.
32. Krishnapriya, S.; Karuna, Y. Pre-trained deep learning models for brain MRI image classification. *Frontiers in Human Neuroscience* **2023**, *17*. <https://doi.org/10.3389/fnhum.2023.1150120>.
33. Morid, M.A.; Borjali, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in Biology and Medicine* **2021**, *128*, 104115. <https://doi.org/10.1016/j.combiomed.2020.104115>.
34. Salehi, A.W.; Khan, S.; Gupta, G.; Alabdullah, B.I.; Almjally, A.; Alsolai, H.; Siddiqui, T.; Mellit, A. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability* **2023**, *15*. <https://doi.org/10.3390/su15075930>.

35. Hossain, M.M.; Hossain, M.M.; Arefin, M.B.; Akhtar, F.; Blake, J. Combining state-of-the-art pre-trained deep learning models: A noble approach for skin cancer detection using max voting ensemble. *Diagnostics* **2023**, *14*, 89.
36. Howard, A.G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* **2017**.
37. Miah, A.S.M.; Shin, J.; Hasan, M.A.M.; Okuyama, Y.; Nobuyoshi, A. Dynamic Hand Gesture Recognition Using Effective Feature Extraction and Attention Based Deep Neural Network. In Proceedings of the 2023 IEEE 16th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc). IEEE, 2023, pp. 241–247.
38. Miah, A.S.M.; Shin, J.; Hasan, M.A.M.; Fujimoto, Y.; Nobuyoshi, A. Skeleton-based hand gesture recognition using geometric features and spatio-temporal deep learning approach. In Proceedings of the 2023 11th European Workshop on Visual Information Processing (EUVIP). IEEE, 2023, pp. 1–6.
39. Miah, A.S.M.; Hasan, M.A.M.; Nishimura, S.; Shin, J. Sign Language Recognition using Graph and General Deep Neural Network Based on Large Scale Dataset. *IEEE Access* **2024**.
40. Miah, A.S.M.; Hasan, M.A.M.; Okuyama, Y.; Tomioka, Y.; Shin, J. Spatial-temporal attention with graph and general neural network-based sign language recognition. *Pattern Analysis and Applications* **2024**, *27*, 37.
41. Pan, J.; Bulat, A.; Tan, F.; Zhu, X.; Dudziak, L.; Li, H.; Tzimiropoulos, G.; Martinez, B. Edgevit: Competing light-weight cnns on mobile devices with vision transformers. In Proceedings of the European Conference on Computer Vision. Springer, 2022, pp. 294–311.
42. Shin, J.; Musa Miah, A.S.; Hasan, M.A.M.; Hirooka, K.; Suzuki, K.; Lee, H.S.; Jang, S.W. Korean Sign Language Recognition Using Transformer-Based Deep Neural Network. *Applied Sciences* **2023**, *13*, 3029.
43. Shin, J.; Miah, A.S.M.; Suzuki, K.; Hirooka, K.; Hasan, M.A.M. Dynamic Korean Sign Language Recognition Using Pose Estimation Based and Attention-Based Neural Network. *IEEE Access* **2023**, *11*, 143501–143513. <https://doi.org/10.1109/ACCESS.2023.3343404>.
44. Mehta, S.; Rastegari, M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178* **2021**.
45. Dai, W.; Liu, R.; Wu, T.; Wang, M.; Yin, J.; Liu, J. Deeply supervised skin lesions diagnosis with stage and branch attention. *IEEE Journal of Biomedical and Health Informatics* **2023**.
46. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *nature* **2017**, *542*, 115–118.
47. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data* **2018**, *5*, 1–9.
48. Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Berking, C.; Steeb, T.; Enk, A.H.; Von Kalle, C. Skin cancer classification using convolutional neural networks: systematic review. *Journal of medical Internet research* **2018**, *20*, e11936.
49. Aljohani, K.; Turki, T. Automatic Classification of Melanoma Skin Cancer with Deep Convolutional Neural Networks. *AI* **2022**, *3*, 512–525.
50. Hasan, M.R.; Fatemi, M.I.; Khan, M.M.; Kaur, M.; Zaguia, A. Comparative Analysis of Skin Cancer (Benign vs. Malignant) Detection Using Convolutional Neural Networks. *Journal of Healthcare Engineering* **2021**, 2021, Article ID 5895156, 17 pages. <https://doi.org/10.1155/2021/5895156>.
51. Balaha, H.; Hassan, A. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Computing and Applications* **2023**, *35*, 815–853. <https://doi.org/10.1007/s00521-022-07762-9>.
52. Han, S.S.; Kim, M.S.; Lim, W.; Park, G.H.; Park, I.; Chang, S.E. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology* **2018**, *138*, 1529–1538.
53. Codella, N.C.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In Proceedings of the 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018, pp. 168–172.
54. Gutman, D.; Codella, N.C.; Celebi, E.; Helba, B.; Marchetti, M.; Mishra, N.; Halpern, A. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1605.01397* **2016**.
55. Barata, C.; Celebi, M.E.; Marques, J.S. Improving dermoscopy image classification using color constancy. *IEEE journal of biomedical and health informatics* **2014**, *19*, 1146–1152.

56. Sahinbas, K.; Catak, F.O. Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images. In *Data science for COVID-19*; Elsevier, 2021; pp. 451–466.
57. Goyal, M.; Knackstedt, T.; Yan, S.; Hassanpour, S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in biology and medicine* **2020**, *127*, 104065.
58. Dildar, M.; Akram, S.; Irfan, M.; Khan, H.U.; Ramzan, M.; Mahmood, A.R.; Alsaiari, S.A.; Saeed, A.H.M.; Al-raddadi, M.O.; Mahnashi, M.H. Skin cancer detection: a review using deep learning techniques. *International Journal of environmental research and public health* **2021**, *18*, 5479.
59. Ahmad, B.; Jun, S.; Palade, V.; You, Q.; Mao, L.; Zhongjie, M. Improving Skin Cancer Classification Using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (TED-GAN). *Diagnostics* **2021**, *11*. <https://doi.org/10.3390/diagnostics11112147>.
60. Rashid, H.; Tanveer, M.A.; Aqeel Khan, H. Skin Lesion Classification Using GAN based Data Augmentation. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 916–919. <https://doi.org/10.1109/EMBC.2019.8857905>.
61. Al-Rasheed, A.; Ksibi, A.; Ayadi, M.; Alzahrani, A.I.A.; Elahi, M.M. An Ensemble of Transfer Learning Models for the Prediction of Skin Lesions with Conditional Generative Adversarial Networks. *Contrast Media & Molecular Imaging* **2023**, 2023, Article ID 5869513, 15 pages. <https://doi.org/10.1155/2023/5869513>.
62. Kebaili, A.; Lapuyade-Lahorgue, J.; Ruan, S. Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review. *Journal of Imaging* **2023**, *9*, 81.
63. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
64. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.
65. Alexey, D. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* **2020**.
66. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 13713–13722.
67. Fan, D.P.; Ji, G.P.; Zhou, T.; Chen, G.; Fu, H.; Shen, J.; Shao, L. Prantet: Parallel reverse attention network for polyp segmentation. In Proceedings of the International conference on medical image computing and computer-assisted intervention. Springer, 2020, pp. 263–273.
68. Lu, X.; Wang, W.; Shen, J.; Crandall, D.; Luo, J. Zero-shot video object segmentation with co-attention siamese networks. *IEEE transactions on pattern analysis and machine intelligence* **2020**, *44*, 2228–2242.
69. Reiß, S.; Seibold, C.; Freytag, A.; Rodner, E.; Stiefelhagen, R. Every annotation counts: Multi-label deep supervision for medical image segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 9532–9542.
70. Shen, J.; Liu, Y.; Dong, X.; Lu, X.; Khan, F.S.; Hoi, S. Distilled Siamese networks for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *44*, 8896–8909.
71. Song, J.; Chen, Y.; Ye, J.; Song, M. Spot-adaptive knowledge distillation. *IEEE Transactions on Image Processing* **2022**, *31*, 3359–3370.
72. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.
73. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018, pp. 116–131.
74. Tan, M. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* **2019**.
75. Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Sandler, M.; Howard, A.; Le, Q.V. Mnasnet: Platform-aware neural architecture search for mobile. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 2820–2828.
76. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314–1324.
77. Suwanwimolkul, S.; Komorita, S. Efficient linear attention for fast and accurate keypoint matching. In Proceedings of the Proceedings of the 2022 international conference on multimedia retrieval, 2022, pp. 330–341.

78. Yang, C.; Wang, Y.; Zhang, J.; Zhang, H.; Wei, Z.; Lin, Z.; Yuille, A. Lite vision transformer with enhanced self-attention. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11998–12008.
79. Gessert, N.; Nielsen, M.; Shaikh, M.; Werner, R.; Schlaefel, A. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX* **2020**, *7*, 100864.
80. Rashid, J.; Ishfaq, M.; Ali, G.; Saeed, M.R.; Hussain, M.; Alkhalifah, T.; Alturise, F.; Samand, N. Skin cancer disease detection using transfer learning technique. *Applied Sciences* **2022**, *12*, 5714.
81. Combalia, M.; Codella, N.C.; Rotemberg, V.; Helba, B.; Vilaplana, V.; Reiter, O.; Carrera, C.; Barreiro, A.; Halpern, A.C.; Puig, S.; et al. Bcn20000: Dermoscopic lesions in the wild. *arXiv preprint arXiv:1908.02288* **2019**.
82. Pacheco, A.G.; Lima, G.R.; Salomao, A.S.; Krohling, B.; Biral, I.P.; de Angelo, G.G.; Alves Jr, F.C.; Esgario, J.G.; Simora, A.C.; Castro, P.B.; et al. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in brief* **2020**, *32*, 106221.
83. OpenCV Team. Image Thresholding — OpenCV-Python Tutorials 1 documentation. https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html, 2024. Accessed: 2025-06-20.
84. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks* **2018**, *106*, 249–259.
85. Smith, M.R.; Martinez, T.; Giraud-Carrier, C. An instance level analysis of data complexity. *Machine learning* **2014**, *95*, 225–256.
86. Li, X.; Liu, T.; Zhang, L.; Alqahtani, F.; Tolba, A. A transformer-BERT integrated model-based automatic conversation method under English context. *IEEE Access* **2024**, *12*, 55757–55767.
87. Ullah, S.; Hassan, N.; Bhatti, N.; Zia, M.; Shin, J. White balancing based improved nighttime image dehazing. *Multimedia Tools and Applications* **2024**, pp. 1–18.
88. Hassan, N.; Ullah, S.; Bhatti, N.; Mahmood, H.; Zia, M. A cascaded approach for image defogging based on physical and enhancement models. *Signal, image and video processing* **2020**, *14*, 867–875.
89. Hassan, N.; Ullah, S.; Bhatti, N.; Mahmood, H.; Zia, M. The Retinex based improved underwater image enhancement. *Multimedia Tools and Applications* **2021**, *80*, 1839–1857.
90. Hassan, N.; Miah, A.S.M.; Suzuki, T.; Shin, J. Gradual Variation-Based Dual-Stream Deep Learning for Spatial Feature Enhancement with Dimensionality Reduction in Early Alzheimer's Disease Detection. *IEEE Access* **2025**.
91. Zuiderveld, K. Contrast limited adaptive histogram equalization. In *Graphics gems IV*; 1994; pp. 474–485.
92. Kim, J.Y.; Kim, L.S.; Hwang, S.H. An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE transactions on circuits and systems for video technology* **2001**, *11*, 475–484.
93. Zhang, L.; Chen, X.; Zhang, J.; Dong, R.; Ma, K. Contrastive deep supervision. In Proceedings of the European Conference on Computer Vision. Springer, 2022, pp. 1–19.
94. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **2019**, *32*.
95. Gollapudi, S. *Learn computer vision using OpenCV*; Springer, 2019.
96. Dozat, T. Incorporating Nesterov Momentum into Adam. In Proceedings of the Proceedings of the 4th International Conference on Learning Representations, Workshop Track, 2016, pp. 1–4.
97. Inthiyaz, S.; Altahan, B.R.; Ahammad, S.H.; Rajesh, V.; Kalangi, R.R.; Smirani, L.K.; Hossain, M.A.; Rashed, A.N.Z. Skin Disease Detection Using Deep Learning. *Advances in Engineering Software* **2023**, *175*, 103361.
98. Alwakid, G.; Gouda, W.; Humayun, M.; Sama, N.U. Melanoma Detection Using Deep Learning-Based Classifications. *Healthcare* **2022**, *10*, 2481.
99. Aljohani, K.; Turki, T. Automatic Classification of Melanoma Skin Cancer with Deep Convolutional Neural Networks. *Ai* **2022**, *3*, 512–525. <https://doi.org/10.3390/ai3040034>.
100. Bechelli, S.; Delhommelle, J. Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images. *Bioengineering* **2022**, *9*, 97. <https://doi.org/10.3390/bioengineering9040097>.
101. Demir, A.; Yilmaz, F.; Kose, O. Early Detection of Skin Cancer Using Deep Learning Architectures: ResNet-101 and Inception-v3. In Proceedings of the Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, October 2019; pp. 1–4. <https://doi.org/10.1109/TIPTEKNO49573.2019.9056683>.
102. kaggle competitions. SIIM-ISIC Melanoma Classification. SIC 2018 - Winners Final 3 Submissions, 2020. SIIM-ISIC Melanoma Classification-Identify melanoma in lesion images, Accessed: 2023-08-22, Available from: <https://www.kaggle.com/competitions/siim-isic-melanoma-classification/discussion/173086w>.

103. Gouda, W.; Sama, N.U.; Al-Waakid, G.; Humayun, M.; Jhanjhi, N.Z. Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare* **2022**, *10*, 1183. <https://doi.org/10.3390/healthcare1011183>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.