Skin Cancer Classification Using Class-Weighted custom EfficientNetB5 Model

A.B. Hoque, 22-46662-1, Department of Computer Science and Engineering, American International University-Bangladesh.

[[1]](#footnote-1)

***Abstract*—** **Skin cancer is one of the top three most dangerous cancers caused by damaged DNA, leading to uncontrolled cell growth. Computerized analysis of skin lesion images for malignancy has been explored, but challenges such as light reflections, varying color lighting, and diverse lesion sizes and shapes complicate accurate analysis. To enhance the precision and support pathologists in early-stage diagnosis, this project proposes a deep learning approach for multi-class skin lesion classification using a customized EfficientNetB5 architecture. To address significant class imbalance in dermatological datasets, a class-weighted function was applied, improving identification of minority classes like melanoma. The model was trained and evaluated on the ISIC-2019 dataset containing 25,531 images across 8 skin lesion categories, achieving a test accuracy of 84.49%, precision of 85.29%, recall of 84.49%, and an F1-score of 84.75%. These results demonstrate that the weighted EfficientNetB5 model is a promising tool for automated skin cancer diagnosis, capable of assisting clinicians in early detection and improving diagnostic workflows. Future work will focus on integrating explainability techniques to enhance model transparency and clinical trust.**

***Index Terms*—** **Skin lesions, deep learning, EfficientNetB5, class imbalance, ISIC-2019**

# I. INTRODUCTION

N

owadays, skin cancer is the most prevalent kind of cancer in people. On rare occasions, it might result in a person dying [1]. Uncontrolled cell growth, fast cell division in one part of the body, invasion of other body tissues, and subsequent body-wide spread are the causes of cancer. Skin cancer can be caused by several unhealthy habits, such as smoking, drinking alcohol, having certain allergies, getting sick, contracting viruses, etc. [2, 3]. This kind of cancer can also be brought on by environmental changes, such as altering the atmosphere. DNA contained within skin cells can be destroyed by the sun's ultraviolet (UV) rays. Additionally, abnormal enlargements of the human body can also result in skin cancer [4]. In the United States, approximately 5.4 million new cases of skin cancer are reported annually [5]. Skin cancer can be broadly divided into two types: melanoma skin cancers, which develop from malfunctioning melanocytes, and non-melanoma skin cancers, which originate from cells originating from the epidermis [5,6]. The majority of skin cancer deaths are caused by melanoma, which is notorious for being aggressive and potentially fatal [6,7]. Like many other forms of cancer, melanoma must be detected early in order to improve patient outcomes and increase treatment options [6,7,8,9]. Early detection not only allows for timely medical intervention but also significantly increases the likelihood of successful treatment, highlighting the importance of advancing diagnostic methods for skin cancer [7,10]. While traditional diagnostic approaches such as visual inspection, dermoscopy, and biopsy remain fundamental, they are often invasive, time-intensive, and heavily reliant on specialized expertise [11]. These challenges, along with escalating healthcare costs, emphasize the critical need for innovative, accurate, and efficient diagnostic technologies to address the growing impact of skin cancer [12]. Artificial intelligence (AI) plays an important role in the early detection of skin cancer, especially DL and machine learning (ML) [13].

In recent times, artificial neural networks (ANNs), an innovative approach within the realm of AI, have gained significant popularity in various domains, including computer vision, digital image processing, and image classification techniques [14]. Within deep learning architectures, convolutional neural networks (CNNs) have established themselves as the cornerstone of medical image analysis [15]. By excelling in the extraction of spatial features, CNNs are particularly effective in identifying cancerous lesions and achieving high levels of accuracy in tasks like skin lesion classification and tumor segmentation n[16]. Vision Transformers (ViTs), a more recent advancement, employ self-attention mechanisms to process image data and have demonstrated their capacity to handle large-scale datasets and extract intricate global patterns [17]. Attallah [18] proposed SCaLiNG, a CAD tool that integrates compact CNNs and Gabor Wavelets to extract spatial, textural, and frequency features. Afza et al. [19] utilized deep feature fusion with extreme learning machines (ELM) for multiclass skin lesion classification. Akram et al. [20] enhanced feature representation by integrating multiple deep models through information-theoretic fusion, while Bibi et al. [21] presented a framework incorporating preprocessing, feature extraction, and classification. To address challenges like data imbalance and model complexity, Ozdemir and Pacal [22] designed a lightweight, mobile-friendly hybrid model. Dillshad et al. [23] proposed an advanced deep learning system for multiclass lesion classification. Naeem et al. introduced SNC\_Net [24], combining CNNs and handcrafted features, and DVFNet [25], which enhances image quality using anisotropic diffusion and extracts features using VGG19 and HOG. Chanda et al. [26] developed DCENSnet, an ensemble of DCNNs with customized dropout for improved feature learning. Brancaccio et al. [27] emphasized the importance of integrating AI tools with human expertise. Pacal et al. [28] improved the Swin Transformer architecture with hybrid shifted window-based multi-head self-attention (HSW-MSA) for better handling of skin lesion intricacies. Cheng et al. [29] combined CNNs with attention mechanisms to capture both local and global features. Attallah [30] also introduced Skin-CAD, an explainable AI system for classifying dermoscopic images into seven subtypes. Riaz et al. [31] conducted a systematic review on federated and transfer learning methods, while Naeem et al. [32] proposed SCDNet, a hybrid model integrating VGG16 and CNNs for classifying melanoma, BCC, and BKL.

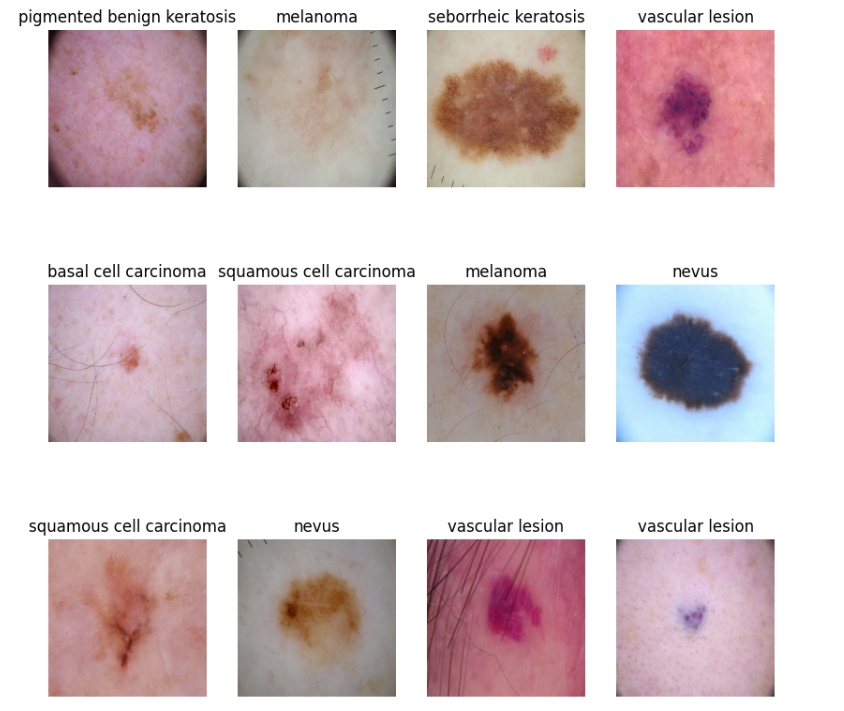
While deep learning models have shown promising results in skin cancer classification, challenges such as dataset imbalance and model generalization still hinder optimal performance. In this context, designing an efficient and balanced architecture becomes crucial for accurate and reliable diagnosis. This study focuses on developing a robust classification system capable of distinguishing between multiple types of skin lesions using deep learning approaches. The main objectives of this study are as follows:

* To develop an automated skin cancer classification system using deep learning techniques.
* To utilize a customized EfficientNetB5 model to achieve a balance between accuracy and computational efficiency.
* To address the issue of class imbalance in the dataset through class-weighted loss functions, ensuring better representation of minority classes.

# II. METHODOLOGY

## A. Dataset and Preprocessing

The ISIC 2019 dataset [36] contains dermoscopic images of eight skin lesion categories, including melanoma, nevus, and basal cell carcinoma. The images vary in number across classes, with benign lesions more frequent. The dataset was split into 80% training and 20% testing sets. All images were resized to 456×456 pixels to match the EfficientNetB5 input requirements and preprocessed using EfficientNet’s standard function. Data augmentation—random flips, brightness, and contrast adjustments—was applied to training images to improve generalization. Class weights were computed to address class imbalance during training. The test images were only resized and normalized without augmentation for evaluation.



**Fig. 1.** Some images with labels of ISIC-2019 dataset.

## B. Proposed model Workflow

The proposed method follows a structured pipeline to classify skin lesion images using a custom EfficientNetB5 architecture. First, the ISIC 2019 dataset images are preprocessed by resizing them to 456×456 pixels and normalized using EfficientNet’s preprocessing function. Data augmentation techniques, such as random flips and brightness adjustments, are applied to enhance model generalization. The dataset is then split into training and testing subsets with an 80:20 ratio. To address the inherent class imbalance, class weights are computed and incorporated during training to reduce bias toward majority classes. Next, a custom EfficientNetB5-based model is built and fine-tuned, where early layers are frozen initially, and later layers are trained with a small learning rate. The model is trained with the weighted data for 10 epochs and a batch size of 16. Finally, the trained model is evaluated on the test set using standard classification metrics including accuracy, precision, recall, F1-score, and confusion matrices to assess its performance and reliability.

A diagram of a model training

AI-generated content may be incorrect.

**Fig. 2.** Proposed Model workflow

## C. Proposed deep convolution neural network (DCNN)

We implemented a custom deep convolutional neural network (DCNN) for multi-class skin lesion classification using the EfficientNetB5 architecture, built with TensorFlow and Keras. EfficientNetB5 was chosen due to its balance of accuracy and computational efficiency, leveraging compound scaling to optimize depth, width, and resolution simultaneously. The model was initialized with pre-trained ImageNet weights to benefit from transfer learning and fine-tuned from layer 300 onward, allowing the network to adapt to skin lesion-specific features while retaining general visual knowledge. On top of the EfficientNetB5 base, a global average pooling layer was added to reduce spatial dimensions, followed by a dense layer with 512 neurons using ReLU activation and L2 regularization (λ=0.005). A dropout layer with a rate of 0.5 was applied to mitigate overfitting. The final classification layer consists of 8 output neurons with softmax activation, corresponding to the 8 disease categories in the ISIC 2019 dataset. The model was compiled using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.0005 and momentum of 0.9. Sparse categorical cross-entropy was used as the loss function to handle integer-labeled multi-class data. The model was trained for 10 epochs with a batch size of 16, using class weights to compensate for data imbalance. This architecture enables the model to extract high-level features from dermoscopic images efficiently, while the added dense and dropout layers enhance the model’s generalization ability and robustness.

A diagram of a machine

AI-generated content may be incorrect.

**Fig. 3.** Proposed EfficientNetB5 Model

# III. MODEL TRAINING

The model was trained on the ISIC 2019 dataset using the preprocessed and augmented training set, with 20% of the data reserved for testing. Training was performed for 10 epochs with a batch size of 16 using the SGD optimizer (learning rate = 0.0005, momentum = 0.9) and sparse categorical cross-entropy as the loss function. The final model achieved a test accuracy of **0.8449**, indicating strong generalization to unseen data.

A graph of a running track

AI-generated content may be incorrect.

**Fig. 4.** Training vs validation accuracy and loss

# IV. EVALUATION METRICES

In addition to accuracy, we will evaluate the model's performance using multiple metrics that provide a comprehensive understanding of its classification ability:

1. Accuracy: Measures the overall correctness of the model's predictions. This is the percentage of correctly classified images out of the total images.

Accuracy =

1. Confusion Matrix: A table used to describe the performance of the classification model. It shows the true positive, true negative, false positive, and false negative rates, which help in understanding how well the model is distinguishing between the different classes.

A diagram of positive values

AI-generated content may be incorrect.

**Fig. 5.** A Confusion Matrix with 2 categories

1. Precision: This metric tells us how many of the predicted positives are actually positive. It’s the ratio of true positives to the total predicted positives.

Precision =

1. Recall (Sensitivity): This metric tells us how many of the actual positives are correctly predicted. It’s the ratio of true positives to the total actual positives.

Recall =

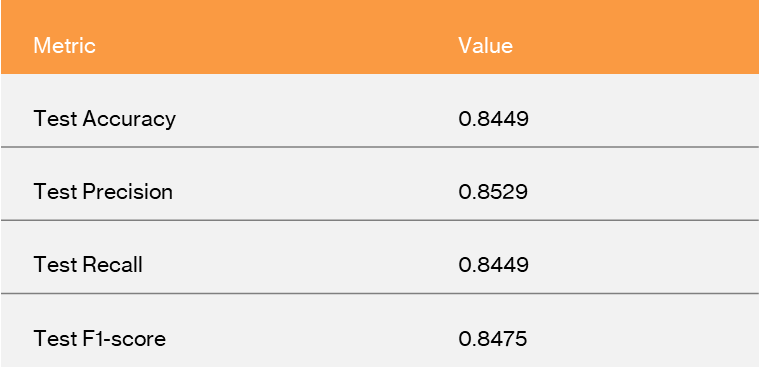
1. F1-Score: The harmonic mean of precision and recall, giving a balance between the two. It is especially useful when dealing with imbalanced datasets.

F1-score =

# V. RESULT AND DISCUSSION

TABLE I

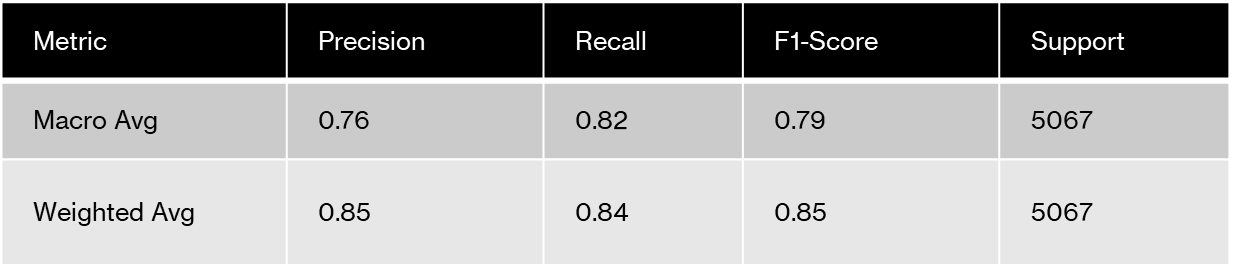
Model Evaluation Metrics on test data



The proposed model achieved a test accuracy of 84.49%, with a precision of 85.29%, recall of 84.49%, and F1-score of 84.75%. These results indicate strong and balanced performance, with the model effectively identifying and classifying skin lesion types with high reliability.

TABLE II

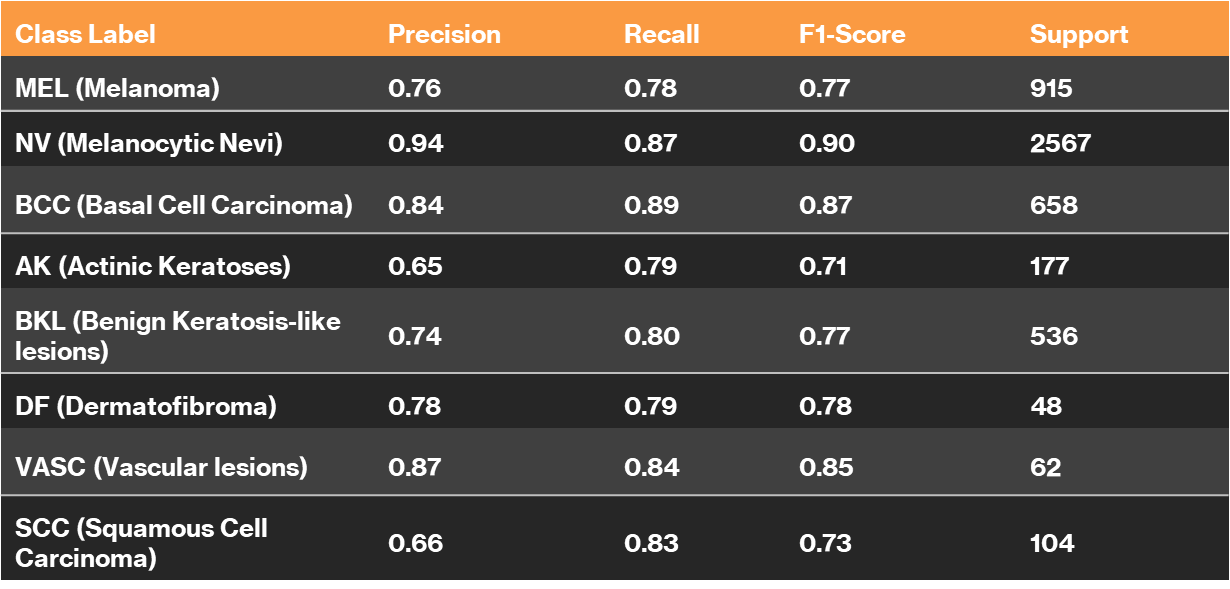
Maro vs Weighted averages



The model performs well overall (weighted F1: 0.85) but shows weaker results for minority classes (macro F1: 0.79). This indicates it handles common cases effectively but struggles with rare ones. Improving class balance through techniques like augmentation or weighted loss could enhance performance across all categories. Based on 5,067 test samples.

TABLE III

Per Class Performance Metrics for Skin-lesion Classification



The per-class performance metrics reveal significant variations in the model's ability to classify different skin lesions. For NV (Melanocytic Nevi), the model achieves strong performance (F1: 0.90), reflecting high precision (0.94) and recall (0.87), likely due to its large sample size (2,567). Similarly, BCC (Basal Cell Carcinoma) and VASC (Vascular lesions) show robust results (F1: 0.87 and 0.85, respectively), with balanced precision and recall. However, the model struggles with rarer classes like AK (Actinic Keratoses) and SCC (Squamous Cell Carcinoma), where precision drops to 0.65 and 0.66, despite moderate recall (0.79 and 0.83). This suggests frequent false positives for these classes.

TABLE IV

Confusion Matrix (without normalization)

A screenshot of a graph

AI-generated content may be incorrect.

TABLE V

Confusion Matrix (with normalization)

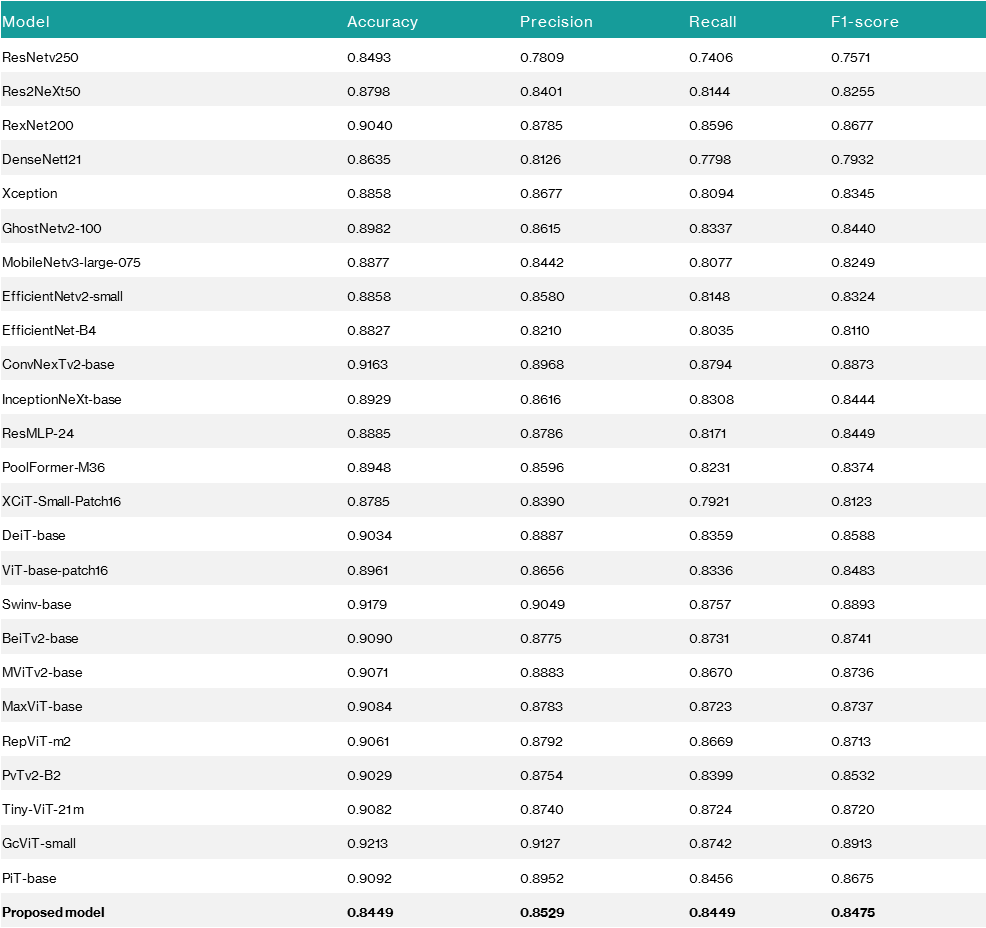
A graph of blue squares

AI-generated content may be incorrect.

The normalized confusion matrix better reveals performance patterns by showing relative error rates across imbalanced classes, while the non-normalized version provides absolute case numbers. Normalization highlights critical issues like the 11% of melanoma (MEL) cases misclassified as NV (nevi) - a dangerous error that appears less prominent in raw counts due to NV's larger sample size. It also better exposes challenges with rare classes (AK, SCC) by showing their misclassification rates (8-9% as BCC) independent of sample size. Though both views are valuable, normalization is superior for evaluating clinical reliability as it equally weights all classes' performance, making it clearer where improvements are most needed, particularly for critical diagnostic distinctions and minority classes.

TABLE VI

Performance Comparison with existing models



The comparative analysis reveals that vision transformer-based models like GeViT-small and ConvNexTv2-base outperform others, achieving the highest accuracy (0.9213) and F1-score (0.8913). The proposed model shows balanced but modest results (F1: 0.8475), lagging behind these top performers. This suggests room for improvement by adopting architectural insights from leading models, such as enhanced self-attention mechanisms or hybrid designs. The dominance of transformer-based approaches highlights their effectiveness in capturing complex patterns for this task. Further optimization of the proposed model could focus on integrating these advanced techniques to boost performance.

VI. FUTURE WORK

In future iterations, several enhancements can be made to further improve model performance and clinical applicability. Data upsampling techniques such as SMOTE or Generative Adversarial Networks (GANs) can be employed to synthetically generate more samples for underrepresented classes, addressing class imbalance more effectively. Incorporating Explainable AI (XAI) methods like Grad-CAM or LIME will help visualize and interpret the model’s predictions, increasing transparency and trust in clinical settings. Additionally, more fine-tuning of deeper layers in the EfficientNetB5 model may allow the network to learn more domain-specific features. Lastly, conducting hyperparameter optimization through grid or random search can help identify optimal values for learning rate, dropout rate, and other parameters to further enhance model accuracy and robustness.

VII. CONCLUSION

The proposed EfficientNetB5-based model demonstrates strong performance in multi-class skin lesion classification, achieving high accuracy, precision, recall, and F1-score. The use of class weights effectively addressed dataset imbalance, contributing to reliable predictions across all categories. Although training was constrained to 30 epochs due to GPU resource limitations, the results indicate that extended training could further enhance model performance. Overall, the model shows significant potential for integration into clinical decision support systems, offering a reliable and automated tool for aiding dermatological diagnosis.

REFERENCES

[1] E. H. Houssein, D. A. Abdelkareem, M. Emam, M. A. Hameed, and M. Younan, "An efficient image segmentation method for skin cancer imaging using improved golden jackal optimization algorithm," *Comput. Methods Programs Biomed.*, vol. 149, p. 106075, 2022.

[2] S. Bibi *et al.*, "MSRNet: Multiclass skin lesion recognition using additional residual block-based fine-tuned deep models, information fusion, and best feature selection," *Diagnostics*, vol. 13, no. 19, p. 3063, 2023.

[3] E. H. Houssein *et al.*, "Liver cancer algorithm: A novel bio-inspired optimizer," *Comput. Methods Programs Biomed.*, vol. 165, p. 107389, 2023.

[4] A. Arshaghi, M. Ashourian, and L. Ghabeli, "Detection of skin cancer image by feature selection methods using new buzzard optimization (BUZO) algorithm," *Trait. Signal*, vol. 37, no. 2, 2020. [Online]. Available: <https://doi.org/10.18280/ts.370204>

[5] N. H. Khan *et al.*, "Skin cancer biology and barriers to treatment: Recent applications of polymeric micro/nanostructures," *J. Adv. Res.*, vol. 36, pp. 223–247, 2022.

[6] R. A. Swerlick, "The melanoma epidemic," *Arch. Dermatol.*, vol. 132, p. 881, 1996.

[7] M. Berwick and A. Halpern, "Melanoma epidemiology," *Curr. Opin. Oncol.*, vol. 9, pp. 178–182, 1997.

[8] J. C. A. Lacson *et al.*, "Skin cancer prevention behaviors, beliefs, distress, and worry among Hispanics in Florida and Puerto Rico," *BMC Public Health*, vol. 23, 2023.

[9] R. S. Werk, J. C. Hill, and J. A. Graber, "Impact of knowledge, self-efficacy, and perceived importance on steps taken toward cancer prevention among college men and women," *J. Cancer Educ.*, vol. 32, pp. 148–154, 2017.

[10] R. Cody and C. Lee, "Behaviors, beliefs, and intentions in skin cancer prevention," *J. Behav. Med.*, vol. 13, pp. 373–389, 1990.

[11] J. W. Kelly, "Melanoma in the elderly: A neglected public health challenge," *Med. J. Aust.*, vol. 169, pp. 403–404, 1998.

[12] R. A. Swerlick, "The melanoma epidemic: More apparent than real?," *Mayo Clin. Proc.*, vol. 72, pp. 559–564, 1997.

[13] N. Ahmad *et al.*, "A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI," *Front. Oncol.*, vol. 13, p. 1151257, 2023.

[14] M. M. Emam, E. H. Houssein, and R. M. Ghoniem, "A modified reptile search algorithm for global optimization and image segmentation: Case study brain MRI images," *Comput. Biol. Med.*, vol. 152, p. 106404, 2023.

[15] I. Pacal and D. Karaboga, "A robust real-time deep learning based automatic polyp detection system," *Comput. Biol. Med.*, vol. 134, 2021.

[16] A. Naeem and T. Anees, "A Multiclassification framework for skin cancer detection by the concatenation of Xception and ResNet101," [Online]. Available: <https://doi.org/10.56979/602/2024>

[17] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *J. Plant Dis. Prot.*, 2024. [Online]. Available: <https://doi.org/10.1007/s41348-024-00896-z>

[18] O. Attallah, “Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and Gabor wavelets,” *Sci. Rep.*, vol. 14, p. 20637, 2024.

[19] F. Afza *et al.*, “Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine,” *Sensors*, vol. 22, p. 799, 2022.

[20] T. Akram *et al.*, “Dermo-optimizer: Skin lesion classification using information-theoretic deep feature fusion and entropy-controlled binary bat optimization,” *Int. J. Imaging Syst. Technol.*, vol. 34, 2024.

[21] S. Bibi *et al.*, “MSRNet: Multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection,” *Diagnostics*, vol. 13, p. 3063, 2023.

[22] B. Ozdemir and I. Pacal, “An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms,” *Results Eng.*, vol. 22, p. 103692, 2024.

[23] V. Dillshad *et al.*, “D2LFS2Net: Multi-class skin lesion diagnosis using deep learning and variance-controlled marine predator optimisation: An application for precision medicine,” *CAAI Trans. Intell. Technol.*, 2023. [Online]. Available: <https://doi.org/10.1049/CIT2.12267>

[24] A. Naeem *et al.*, “SNC\_Net: Skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images,” *Mathematics*, vol. 12, p. 1030, 2024.

[25] A. Naeem, T. Anees, and DVFNet, “A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images,” *PLoS One*, vol. 19, e0297667, 2024.

[26] D. Chanda *et al.*, “A new deep convolutional ensemble network for skin cancer classification,” *Biomed. Signal Process. Control*, vol. 89, p. 105757, 2024.

[27] G. Brancaccio *et al.*, “Artificial intelligence in skin cancer diagnosis: A reality check,” *J. Invest. Dermatol.*, vol. 144, pp. 492–499, 2024.

[28] I. Pacal, M. Alaftekin, and F. D. Zengul, “Enhancing skin cancer diagnosis using Swin Transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP,” *J. Imaging Inf. Med.*, pp. 1–19, 2024. [Online]. Available: <https://doi.org/10.1007/S10278-024-01140-8>

[29] H. Cheng, J. Lian, and W. Jiao, “Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism,” *Sci. Rep.*, vol. 14, pp. 1–13, 2024.

[30] O. Attallah, “Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning,” *Comput. Biol. Med.*, vol. 178, p. 108798, 2024.

[31] S. Riaz, A. Naeem, H. Malik, R. A. Naqvi, and W. K. Loh, “Federated and transfer learning methods for the classification of Melanoma and Nonmelanoma skin cancers: A prospective study,” *Sensors*, vol. 23, p. 8457, 2023.

[32] A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S. W. Lee, “SCDNet: A deep learning-based framework for the multiclassification of skin cancer using dermoscopy images,” *Sensors*, vol. 22, p. 5652, 2022.

[33] I. Goodfellow, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, pp. 2234–2242, Oct. 2014.

[34] I. S. A. Abdelhalim, M. F. Mohamed, and Y. B. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," *Expert Syst. Appl.*, vol. 165, Mar. 2021.

1. [↑](#footnote-ref-1)