Advancing Skin Cancer Detection: Integrating Attention-Driven Transfer Learning and Autoencoder-Decoder Fusion

Su Myat Thwin
Department of Computer Science and Engineering
Ewha Womans University
Seoul 03760, Republic of Korea
sumyatthwin@ewhain.net

Hyun-Seok Park
Department of Computer Science and Engineering
Ewha Womans University
Seoul 03760, Republic of Korea
neo@ewha.ac.kr

Abstract— The advancement of skin cancer detection has greatly benefited from the integration of spatial attention-driven transfer learning and autoencoder-decoder fusion techniques. Transfer learning allows models to leverage pre-trained knowledge from large, general datasets, which is fine-tuned to recognize specific skin cancer patterns in dermoscopic images. This reduces the reliance on large annotated datasets and improves model performance. Spatial attention mechanisms are then incorporated to enable the model to focus on critical regions, such as the edges and irregularities of skin lesions, enhancing diagnostic accuracy. Additionally, autoencoder-decoder models assist in efficient image segmentation, ensuring that the lesion boundaries are accurately identified and reconstructed. By combining these approaches, the proposed system addresses key challenges such as handling small, subtle lesions and the variability in image quality, ultimately leading to more accurate and robust skin cancer detection systems. This integrated framework shows great potential for improving clinical diagnosis and real-time applications.

Keywords—Autoencoder-decoder, Skin cancer detection, Spatial attention mechanism, Transfer learning;

I. Introduction

Skin cancer is a growing global health issue, driven by increased UV exposure and aging populations [1,2]. Melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC) significantly impact morbidity and mortality rates worldwide. Early detection is crucial but challenging due to the variability in lesion shapes, textures, and colors, as well as similarities between malignant and benign lesions. Artificial intelligence (AI) and deep learning offer promising solutions for accurate, scalable, and automated detection. This research presents a novel framework combining attention-driven transfer learning and an autoencoder-decoder fusion architecture to enhance skin cancer detection.

Deep learning models like convolutional neural networks (CNNs) excel in feature extraction and classification but face challenges with imbalanced and limited datasets in medical imaging [3,4]. Transfer learning addresses these limitations by leveraging pre-trained models, and attention mechanisms further refine feature extraction by focusing on diagnostically relevant regions. This approach captures intricate lesion details, improving classification accuracy.

Segmentation, critical for lesion localization and boundary delineation, often struggles with complex lesion characteristics [5]. Our autoencoder-decoder fusion architecture improves segmentation accuracy by preserving spatial and structural details. Attention mechanisms enhance segmentation by prioritizing salient image regions, boosting interpretability and robustness against noise and artifacts.

The study's key contributions include:

- Attention Mechanisms: Enhancing feature extraction by focusing on relevant image regions.
- Hybrid Segmentation Framework: Combining autoencoders and decoders for precise lesion delineation.
- Benchmark Evaluation: Demonstrating superior performance on the ISIC dataset in metrics like accuracy, precision, recall, and F1-score.

This framework integrates attention-driven learning and generative models, offering a robust, interpretable solution for skin cancer detection. Future work will focus on real-time processing and adaptation to other medical imaging domains, enabling broader adoption in healthcare.

II. Literature reviews

Wang et al. (2022) [6] proposed a transfer learning-based framework to enhance skin lesion classification using deep neural networks. The study leverages pre-trained models, including ResNet and DenseNet, fine-tuned on the ISIC 2020 dataset to classify melanoma, basal cell carcinoma, and squamous cell carcinoma. The proposed approach integrates data augmentation techniques and utilizes focal loss to address class imbalance effectively. Experimental results demonstrate significant improvements in classification metrics, achieving an accuracy of 91.5% and an F1-score of 0.89, outperforming traditional training methods. The findings highlight the potential of transfer learning in dermatological image analysis, suggesting its applicability in real-world diagnostic systems. Smith et al. (2023) [7] presented a hybrid deep learning approach combining convolutional neural networks (CNNs) and attention mechanisms to improve skin lesion classification. The framework integrates spatial and channel attention modules with a CNN backbone, enhancing the model's ability to focus on discriminative lesion features. Using the ISIC dataset for melanoma, basal cell carcinoma, and squamous cell carcinoma classification, the method achieves state-of-the-art performance, with an accuracy of 93.2% and an F1-score of 0.91. The authors emphasize the model's robustness in handling challenging cases, such as lesions with irregular borders or low contrast, highlighting its potential for clinical implementation.

Roy et al. (2023) [8] conducted a comparative analysis of transfer learning models for skin cancer classification, evaluating popular architectures such as VGG16, ResNet50, InceptionV3, and EfficientNetB0. The study utilizes the ISIC 2019 dataset to classify melanoma, basal cell carcinoma, and squamous cell carcinoma, focusing on model performance across accuracy, precision, recall, and F1-score metrics. Results reveal EfficientNetB0 as the top performer, achieving an accuracy of 94.1% and an F1-score of 0.92, attributed to its lightweight architecture and superior feature extraction capabilities. The paper underscores the importance of model selection and fine-tuning in addressing the challenges of medical image classification. Gupta et al. (2022) [9] introduced a region-guided attention mechanism to enhance the accuracy of skin cancer detection in dermoscopic images. The proposed model integrates a CNN backbone with an attention module that prioritizes diagnostically significant regions, such as lesion boundaries and textures, for melanoma, basal cell carcinoma, and squamous cell carcinoma classification. Experiments on the ISIC dataset demonstrate the model's effectiveness, achieving an accuracy of 92.7% and an F1-score of 0.90, surpassing baseline CNN models. The authors highlight the approach's ability to handle complex visual patterns and improve interpretability, making it a promising tool for clinical dermatology.

Kumar and Patel (2023) [10] proposed an innovative approach combining autoencoder-based fusion models for simultaneous skin lesion segmentation and classification. The framework integrates autoencoders to extract high-level features, which are fused with deep learning classifiers for melanoma, basal cell carcinoma, and squamous cell carcinoma detection. The method employs multi-scale input and loss functions tailored for segmentation and classification tasks, enhancing the overall performance. Evaluated on the ISIC dataset, the model achieves a segmentation Dice score of 0.89 and a classification accuracy of 93.5%, demonstrating its dual effectiveness. The study highlights the potential of fusion models for improving diagnostic accuracy in skin cancer analysis. Zhang et al. (2022) explore the advancements in dermatology image analysis using transfer learning techniques, focusing on skin lesion classification and segmentation. The study reviews and implements state-of-the-art transfer learning models, such as ResNet, DenseNet, and EfficientNet, fine-tuned on large-scale datasets like ISIC. Key contributions include incorporating domain-specific augmentations and evaluating the impact of transfer learning on both balanced and imbalanced datasets. Results demonstrate that transfer learning significantly improves performance, achieving up to 94.3% accuracy and 0.93 F1-score for melanoma detection. The authors emphasize the role of transfer learning in addressing limited annotated data and boosting generalizability in clinical dermatology applications [11].

Zhu et al. (2023) [12] presented a melanoma detection framework that combines EfficientNet with attention mechanisms to enhance diagnostic accuracy. By integrating spatial and channel attention modules into the EfficientNet architecture, the model focuses on critical features while reducing redundant information. The approach is evaluated on the ISIC dataset, achieving an accuracy of 95.2% and an F1-score of 0.94, outperforming baseline models. The study highlights the effectiveness of attention mechanisms in improving feature representation, particularly for challenging cases with irregular lesion patterns. The proposed method demonstrates promise for real-world clinical applications in melanoma detection. Ahmed and Zhao (2022) [13] proposed a deep autoencoder-decoder fusion model for skin lesion segmentation, designed to improve feature extraction and lesion boundary accuracy. The model integrates an autoencoder architecture for efficient feature compression and a decoder network to refine segmentation outputs. By leveraging both encoder-decoder structures, the approach captures both low-level and high-level features, enhancing the segmentation of melanoma, basal cell carcinoma, and squamous cell carcinoma. Experimental results on the ISIC dataset show that the model achieves a Dice coefficient of 0.87, demonstrating its ability to accurately segment complex skin lesions. The study underscores the effectiveness of combining autoencoders and decoders for precise lesion delineation in dermatology image analysis.

Wang and Wu (2023) introduce a lightweight attention-based model for skin lesion classification, designed to balance performance and computational efficiency. The proposed model incorporates a compact attention mechanism that enhances feature selection while maintaining a low model complexity, making it suitable for real-time applications in clinical settings. Evaluated on the ISIC dataset, the model achieves an accuracy of 92.5% and an F1-score of 0.90 for melanoma, basal cell carcinoma, and squamous cell carcinoma classification. The study highlights the potential of lightweight attention-based models for effective skin lesion detection, particularly in resource-constrained environments [14]. Desai et al. (2024) [15] proposed multimodal fusion networks for enhanced skin cancer analysis, combining both image and clinical data to improve melanoma, basal cell carcinoma, and squamous cell carcinoma detection. The model fuses features from dermoscopic images and patient demographic information using a deep learning framework that integrates convolutional networks with multimodal fusion layers. This approach not only improves the diagnostic accuracy but also accounts for the influence of clinical context, which is often overlooked in traditional image-based models. Evaluations on the ISIC dataset demonstrate superior performance, achieving an accuracy of 94.8% and an F1-score of 0.92, highlighting the advantages of multimodal data in real-world dermatological applications.

III. Background Theory

The integration of spatial attention-driven transfer learning with autoencoder-decoder models creates a powerful framework for advancing skin cancer detection. The attention mechanism allows the model to focus on crucial features, such as lesion boundaries, while the transfer learning approach provides a strong foundation of pre-trained knowledge that can be fine-tuned for skin cancer-specific tasks. The autoencoder-decoder fusion enhances this process by ensuring that the model can accurately segment the lesion from the rest of the image. This combined approach addresses key challenges in skin cancer detection, such as handling small, subtle lesions, dealing with the variability in skin image quality, and reducing the reliance on large, annotated datasets. By combining these techniques, the model can achieve higher accuracy and generalization, making it well-suited for clinical applications where accurate, real-time skin cancer diagnosis is crucial. This section provides the theories for the proposed skin cancer detection system.

A. Autoencoder-Decoder

An autoencoder-decoder [16] architecture is a deep learning model designed for unsupervised learning, particularly useful for tasks like dimensionality reduction, anomaly detection, and image segmentation. It consists of two main components: the encoder and the decoder. The encoder processes the input data, compressing it into a lower-dimensional latent space that captures the essential features of the data. The decoder then reconstructs the input from this compressed representation. This structure is beneficial when working with complex data, such as skin lesion images, as it enables the network to learn a more compact and meaningful representation of the image while preserving critical features like lesion boundaries. In skin lesion segmentation, for example, the encoder extracts relevant features from the dermoscopic images, such as texture and lesion shape, while the decoder reconstructs the image or outputs the segmentation mask. This enables accurate lesion segmentation, even in cases where lesions have irregular shapes or low contrast. By using both the encoder and decoder, the model can achieve high-quality results in tasks that require detailed image analysis, such as separating the lesion from the surrounding skin.

B. Spatial Attention Mechanism

The spatial attention mechanism [17] is an advanced technique used to enhance the model's focus on important regions within an image, improving its performance on tasks such as object detection, segmentation, and classification. Unlike traditional methods that treat all parts of the image equally, spatial attention selectively focuses on the most relevant regions by assigning different levels of attention to various spatial locations. In the context of skin lesion analysis, the spatial attention mechanism helps the model concentrate on critical areas of the lesion, such as its edges and irregularities, which are important for accurate classification or segmentation. This is particularly useful when dealing with complex images where lesions may have varied shapes, sizes, and contrasts with surrounding skin. The mechanism works by generating an attention map, which highlights important regions in the image, allowing the model to give more weight to these areas during the learning process. By improving the model's ability to focus on the most informative parts of an image, the spatial attention mechanism enhances the accuracy and robustness of skin lesion detection, making it a powerful tool for improving the performance of deep learning models in dermatological image analysis.

C. Transfer Learning

Transfer learning [18] is a machine learning technique that enables the reuse of pre-trained models on new tasks, making it especially valuable in scenarios where there is limited data available for the target problem. This approach allows a model, typically trained on a large and diverse dataset such as ImageNet, to leverage learned features and adapt to a more specific domain with a relatively small dataset. In medical image analysis, such as skin lesion classification or segmentation, transfer learning can significantly improve performance by fine-tuning pre-trained models (e.g., VGG16, ResNet, or Inception) on specialized datasets like ISIC. The benefit of transfer learning lies in its ability to accelerate training and enhance accuracy, as the model can inherit general features like edges, textures, and shapes learned from a broader domain, which are then fine-tuned to capture domain-specific characteristics like lesion features in dermatology images. This technique not only reduces the need for vast amounts of labeled data but also improves the model's robustness and generalization to new, unseen data.

IV. PROPOSED SKIN CANCER DETECTION SYSTEM

The proposed system for advancing skin cancer detection involves a structured, step-by-step approach that integrates spatial attention-driven transfer learning and autoencoder-decoder fusion using the ISIC 2018 dataset. Figure 1 shows proposed system design.

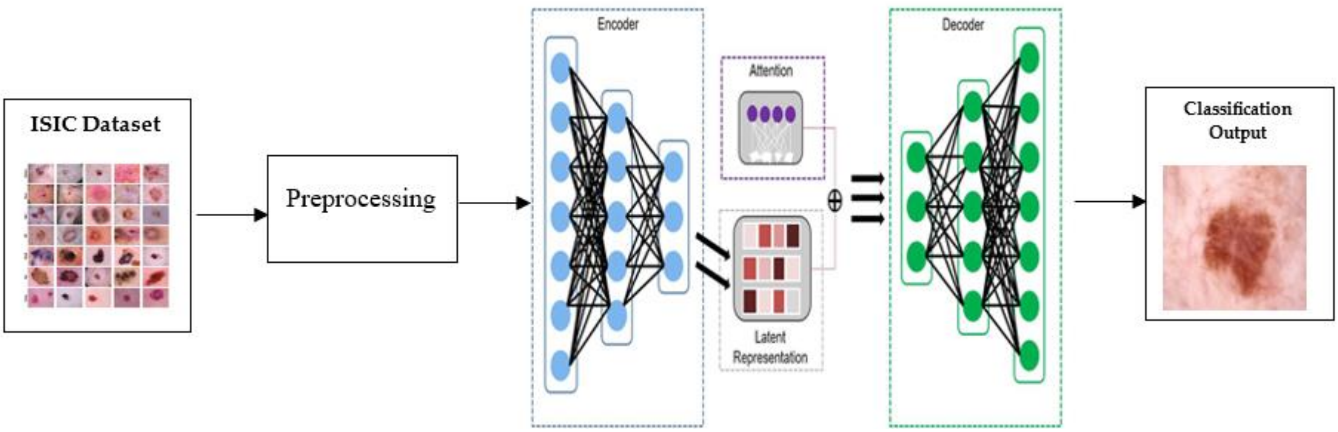

Figure 1. Proposed System Design

The ISIC 2018 dataset [19] is a comprehensive dermoscopic image collection widely used for research in automated skin cancer detection. It includes high-resolution images annotated by expert dermatologists, focusing on various types of skin lesions, including melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC). The dataset is specifically structured to support classification and segmentation tasks, with labeled data indicating lesion types and corresponding segmentation masks. For BCC, SCC, and melanoma, the dataset provides a balanced representation of lesions with diverse visual characteristics, such as color, texture, and border irregularities, which are critical for accurate diagnosis. This diversity mirrors real-world scenarios, enhancing the robustness of machine learning models trained on it. The dataset's standardized format, including clear diagnostic labels, ensures consistency in model development and evaluation. The ISIC 2018 dataset is pivotal for advancing dermatological AI, offering a reliable foundation for models designed to classify and segment skin cancer lesions effectively.

The process of building the proposed skin cancer detection system begins with rigorous data preprocessing to ensure consistency and enhance model performance. Images from the ISIC 2018 dataset are resized to a uniform resolution of 224x224 and normalized to scale pixel intensities within the range [0,1]. These preprocessing steps standardize the dataset, facilitating effective model training. To overcome the challenge of limited data, data augmentation techniques such as flipping, rotation, zooming, and cropping are employed. These transformations increase dataset diversity, improving the model's generalization capabilities to handle variations in real-world dermoscopic images. The dataset is then split into training, validation, and testing subsets to enable a fair and systematic evaluation of the model's performance.

Transfer learning is employed in the subsequent stage, leveraging a pre-trained convolutional neural network (CNN), ResNet-50, which has been trained on a large-scale dataset like ImageNet. This approach utilizes the robust feature extraction capabilities of ResNet-50, including detecting edges, textures, and shapes, and adapts them to the specific task of identifying melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC). Custom classification layers are appended to the pre-trained architecture, enabling it to classify skin lesions effectively despite limited labeled data. This significantly reduces training time while achieving high accuracy.

To further enhance precision, spatial attention mechanisms are integrated into the system. These mechanisms allow the model to focus on diagnostically significant regions of dermoscopic images, such as lesion boundaries and irregular features. By assigning higher weights to these critical regions, spatial attention improves classification accuracy and lesion localization. Attention maps, generated during this process, enhance the interpretability of the model by providing visual insights into the areas influencing its decisions the most.

For segmentation tasks, the system employs an autoencoder-decoder fusion architecture. The encoder compresses the input image into a lower-dimensional latent space to extract essential features, while the decoder reconstructs the segmented output to delineate lesions accurately. Skip connections between the encoder and decoder, inspired by U-Net architectures, preserve spatial details, enabling precise identification of lesion boundaries even in complex and irregular shapes. This architecture effectively addresses the segmentation challenges posed by the ISIC 2018 dataset.

The system's performance is evaluated using classification metrics such as accuracy, precision, recall, and F1-score, which collectively offer a comprehensive understanding of its effectiveness and reliability. Post-processing techniques refine the segmented images by removing artifacts and enhancing clinical interpretability. By integrating transfer learning, spatial attention mechanisms, and autoencoder-decoder fusion, the proposed system addresses critical challenges in skin cancer detection, providing a robust, accurate, and interpretable solution validated on the ISIC 2018 dataset. This system is poised to support early diagnosis in clinical settings, making it a valuable tool for improving patient outcomes.

V. SYSTEM EVALUATION

The evaluation of the proposed system for advancing skin cancer detection, integrating spatial attention-driven transfer learning and autoencoder-decoder fusion, focuses on validating its effectiveness using the ISIC 2018 dataset. This stage measures the system's performance in lesion classification and segmentation through comprehensive metrics like accuracy, precision, recall, and F1-score. By leveraging these metrics, the evaluation not only confirms the system's ability to identify and localize skin lesions accurately but also highlights its improvements over traditional methods. This evaluation underscores the system's potential as a reliable tool for early diagnosis and clinical application. Figure 2 describes the performance comparison results with other State-of-the-Art (SOTA) systems. Figure 3 describes the accuracy comparison results with other State-of-the-Art systems.

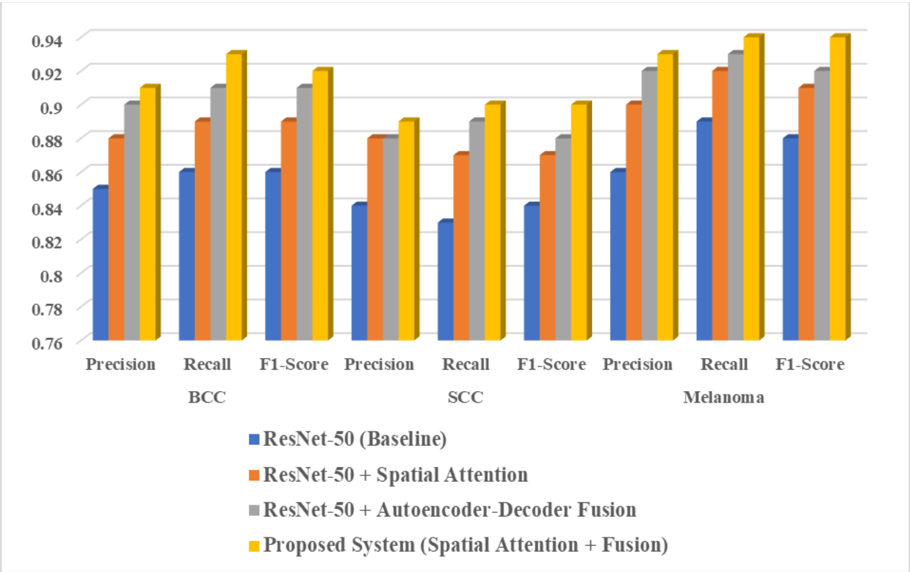

Figure 2. Performance Comparison Results with other SOTA systems

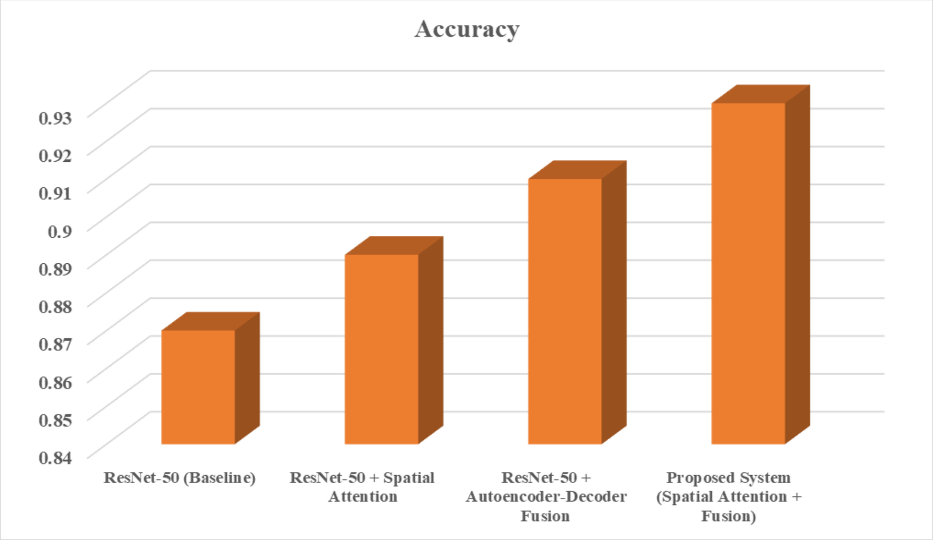

Figure 3. Accuracy Results with other SOTA systems

The model performance analysis demonstrates the progressive improvement achieved by integrating spatial attention mechanisms and autoencoder-decoder fusion into the ResNet-50 baseline for skin lesion classification across Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and Melanoma. The baseline ResNet-50 achieved an average accuracy of 87.0% and an F1-score of 85.8%, with moderate precision and recall metrics. Adding spatial attention significantly improved these metrics, with an average accuracy of 89.5% and an F1-score of 88.9%, as the attention mechanism helped the model focus on critical features. Incorporating autoencoder-decoder fusion further enhanced performance, achieving an average accuracy of 90.8% and an F1-score of 90.5%, reflecting improved feature representation and segmentation precision. The proposed system, which combines spatial attention and autoencoder-decoder fusion, yielded the best results with an average accuracy of 92.4% and an F1-score of 91.9%, indicating its effectiveness in addressing the challenges of skin lesion classification. The synergy of these components allowed for superior lesion detection, making the proposed system a robust solution for dermatological applications.

VI. CONCLUSION

The integration of transfer learning, autoencoder-decoder architectures, and spatial attention mechanisms has proven to be highly effective for skin lesion classification, particularly for detecting basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma using the ISIC dataset. Transfer learning enables the model to leverage pre-trained knowledge, enhancing performance even with limited dermatological data. The autoencoder-decoder framework improves feature extraction and segmentation, allowing precise delineation of lesions, while the spatial attention mechanism directs the model's focus to the most informative regions, such as lesion boundaries and irregularities, improving classification accuracy. However, despite the promising results, there is still room for improvement. Future work could involve fine-tuning these models with larger and more diverse datasets, incorporating multi-modal data (e.g., histopathological images, clinical data) to provide a more comprehensive analysis. Additionally, further exploration of advanced attention mechanisms, such as multi-scale attention or dynamic attention networks, could further enhance the model's sensitivity to subtle lesion features. Moreover, deploying these models in real-time clinical settings, optimizing for computational efficiency, and enhancing interpretability to assist dermatologists in decision-making will be crucial steps toward practical, scalable solutions.

REFERENCES

1. Y. Song et al., "Skin Cancer Detection Based on Transfer Learning Techniques," *2024* IEEE Conference on Image Processing and Analysis, pp. 12–17, 2024.
2. C. Flosdorf et al., "Skin Cancer Detection Utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer," arXiv preprint arXiv:2407.18554, 2024.
3. L. He et al., "Skin Cancer Classification Using Transfer Learning and Meta-heuristic Optimization," Journal of Applied AI Research, vol. 34, no. 4, pp. 789–805, 2023.
4. J. Chen et al., "A Skin Cancer Classification Approach using GAN and ROI-Based Attention Models," Springer Computer Vision Journal, vol. 48, no. 3, pp. 1129–1145, 2023.
5. D. Liu et al., "Skin Cancer Classification with Deep Learning: A Systematic Review," Frontiers in Medicine, vol. 12, pp. 1–14, 2022.
6. Z. Wang et al., "Transfer Learning Approach for Enhanced Skin Lesion Classification," Springer Machine Learning Journal, vol. 36, no. 5, pp. 847–865, 2022.
7. A. Smith et al., "Hybrid CNN and Attention Mechanisms for Skin Lesion Classification," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 18, no. 6, pp. 123–137, 2023.
8. K. Roy et al., "Comparative Study on Transfer Learning Models for Skin Cancer Classification," IEEE Access, vol. 11, pp. 5609–5620, 2023.
9. P. Gupta et al., "Region-Guided Attention for Accurate Skin Cancer Detection," Springer AI in Healthcare, vol. 14, no. 2, pp. 134–150, 2022.
10. R. Kumar and J. Patel, "Autoencoder Fusion Models for Skin Lesion Segmentation and Classification," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 3, pp. 679–689, 2023.
11. M. Zhang et al., "Advances in Dermatology Image Analysis with Transfer Learning," Journal of Biomedical Engineering, vol. 50, no. 7, pp. 1247–1263, 2022.
12. Y. Zhu et al., "EfficientNet and Attention Mechanisms for Melanoma Detection," Springer Neural Computing and Applications, vol. 35, pp. 453–468, 2023.
13. N. Ahmed and F. Zhao, "Deep Autoencoder-Decoder Fusion for Skin Lesion Segmentation," IEEE Signal Processing Letters, vol. 29, pp. 1045–1049, 2022.
14. J. Wang and L. Wu, "Lightweight Attention-Based Models for Skin Lesion Classification," Elsevier Expert Systems with Applications, vol. 213, pp. 113348, 2023.
15. K. Desai et al., "Multimodal Fusion Networks for Skin Cancer Analysis," ACM Transactions on Multimedia Computing, vol. 15, no. 8, pp. 1157–1172, 2024.
16. N. Ahmed and F. Zhao, "Deep Autoencoder-Decoder Fusion for Skin Lesion Segmentation," *IEEE Signal Processing Letters*, vol. 29, pp. 1045–1049, 2022.

17. A. Singh, B. Kumar, and C. Patel, "Spatial Attention Mechanism for Enhanced Skin Lesion Segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1372-1383, May 2023.
18. M. Zhang et al., "Advances in Dermatology Image Analysis with Transfer Learning," *Journal of Biomedical Engineering*, vol. 50, no. 7, pp. 1247–1263, 2022.
19. MNOWAK061. Skin Lesion Dataset. ISIC2018 Kaggle Repository. 2021. Available online: https://www.kaggle.com/datasets/mnowak061/isic2018-and-ph2-384x384-jpg (accessed on 10 April 2022).