Received: 22 October 2025 | Revised 18 November 2025 | Accepted 27 December 2025 | Page 19- 27

Research Article
Journal of Research and Review in Science
Volume 12, 19-27, Issue 2 December 2025
ORIGINAL RESEARCH

IMPROVING SKIN LESION SEGMENTATION IN DERMOSCOPY IMAGES USING DATA AUGMENTATION AND PRE-TRAINED MODEL

Kehinde Sotonwa[1], Tomiloba Olowo[1], Hanat Raji- Lawal[1], Adedoyin Odumabo[2], Folasade Okikiola[3], Idris Aremu[2], Mariam Aliyu[1], Ogunyemi Oluwapelumi[1]

[1]Department of Computer Science, Faculty of Computing and Information Technology, Lagos State University, Nigeria

[2]Department of Computer Sciences, College of Basic Sciences, Lagos State University of Science and Technology, Nigeria

[3]Department of Computer Technology, School of Technology, Yaba College of Technology, Nigeria

Correspondence
Kehinde Adebola Sotonwa, Department of Computer Science, Faculty of Computing and Information Technology, Lagos State University, Nigeria.
Email: kehinde.sotonwa@lasu.edu.ng

Abstract:
Introduction: Skin cancer incidence is increasing globally, making early and accurate diagnosis essential. Dermoscopy improves detection but is limited by clinician variability. Automated segmentation using deep learning offers a promising alternative, though challenges such as image variability and artifacts remain.
Aims: To develop and evaluate an improved UNet++-based model for accurate skin lesion segmentation and classification in dermoscopic images.
Materials and Methods: The study utilised the ISIC 2018 dataset comprising 2,694 dermoscopic images with corresponding expert-annotated segmentation masks. A UNet++ architecture with an ImageNet-pretrained EfficientNet-B5 encoder was implemented. Data augmentation techniques, including geometric and photometric transformations, were applied to improve generalisation. A composite loss function combining Dice loss (50%), binary cross-entropy (30%), and focal loss (20%) was used to address class imbalance and improve boundary delineation. Model performance was evaluated using sensitivity, specificity, accuracy, Dice coefficient (DC), Intersection over Union (IoU), F1-score, and area under the curve (AUC).
Results: The model achieved classification sensitivity of 0.8900, specificity of 0.9200, accuracy of 0.9500, F1-score of 0.8870, and AUC of 0.9278. For segmentation, it achieved a Dice coefficient of 0.8870 and an IoU of 0.8200. The model outperformed U-Net and showed consistent improvements over O-Net across key metrics.
Conclusion: The proposed UNet++ framework improves segmentation accuracy and classification performance, demonstrating strong potential for clinical application in automated skin cancer diagnosis.
To Keywords: Skin cancer, Dermoscopy, Segmentation, UNet++, EfficientNet-B5, Deep learning

All co-authors agreed to have their names listed as authors.

1. INTRODUCTION

Skin cancer is one of the most serious health problems worldwide. Every year, there are about 2 to 3 million cases of diagnosed non-melanoma skin cancer and about 132,000 melanomas worldwide [1]. The World Health Organisation estimates that by 2040, they may increase by 50%, due to increased exposure to UV radiation, an ageing population, ozone layer depletion, and changes in leisure activities [2]. Early detection is of paramount importance: while the 5-year survival rate for localised melanoma is over 95%, it decreases to less than 25% once the tumour has spread to distant parts of the body [3]. These striking contrasts underpin the need for the development of accurate and available diagnostic modalities [4].

Current diagnostic methods rely significantly on visual examination by dermatologists, often aided with dermoscopy, a non-invasive technique offering 10x magnification through polarised light [5]. Though dermoscopy enhances diagnostic performance compared to the naked eye by 15-20%, this still varies appreciably with the expertise of the clinician, generally settling between 75-84% even within experienced practitioners [5]. These limitations are further compounded by substantial inter-observer variability -(κ = 0.4-0.6)-and limited access in under-resourced settings [6]. The financial implications are equally staggering, with annual U.S. treatment costs surpassing $8.1 billion, $3.3 billion of which relates specifically to melanoma management [4]. These challenges have sparked considerable research into CAD systems designed to support clinical decision-making and increase access to quality screening [7].

Deep learning has revolutionised medical image analysis, including dermatology. CNNs have exhibited outstanding abilities in pattern recognition and feature extraction. This led to intensive studies in different CNN architectures on skin lesion analysis [8]. The transformation from conventional image processing methodologies, such as Thresholding, edge detection, and region-based techniques, to advanced deep learning approaches has been rapid and transformative [8]. FCNs brought the notion of end-to-end dense prediction. The U-Net architecture with symmetric encoder-decoder structure, along with skip connections, became the golden standard for medical image segmentation, allowing Dice coefficients of 0.83-0.85 on benchmark datasets [9]. Later variants included the U-Net++, which introduced nested skip connections for multiscale feature fusion. The DeepLab family of models incorporated atrous convolution along with spatial pyramid pooling to capture more multi-scale context information [5].

Recently, the transformer-based architecture has been accepted as an alternative to CNNs; early works focused on a particular instance of skin lesion segmentation and introduced recurrent attentional convolutional networks. Their approach emphasised attention to dealing with certain dermoscopic images clinically and referred to these images to catch long-range dependencies. Their model obtained a Dice coefficient of 0.85. Their model demonstrated increased robustness against image quality fluctuations relative to previous CNN-based approaches. First, drastic levels of attention on certain images increase the model's computational complexity and increase its memory and processing requirements. This complexity is also slowing the model down during training and inference. Second, the sensitivity of the model during training was also explained. Substantial performance loss has been demonstrated with images taken in fluctuations of the clinical setting and during different lighting scenarios. Their models are known to work poorly with their training conditions during turbulence in the acquisition techniques. Third, it was remarked that the difference between model experimental performance and model usability in practice, particularly long duration processing times, is incompatible with real diagnostic workflows. Lastly, they pointed out that the 'black box' nature of deep learning models, i.e. opacity, could hinder usability and consequently trust in the clinical context [9].

In this study, we attempt to expand upon what was proposed by designing an updated deep learning-based model that addresses the problems described while trying to retain (or improve on) the same level of accuracy with respect to segmentation. To avoid using expensive computational transformer encoders, we take an UNet++ architecture with an EfficientNet-B5 encoder, which happens to strike the right balance of feature extraction and computational concerns. The EfficientNet models were designed specifically to obtain better accuracy with far fewer parameters due to the compound scaling of network depth, width, and resolution [10]. The challenge of generalisation described by [9] has been tackled by effective and diverse geometric and photometric augmentation, the simulation of occlusion, and their

implementation in the model to ensure the widest variety of adaptability of the model to distinctive imaging parameters.

The integration of Dice Loss, Binary Cross-Entropy, and Focal Loss into a composite loss function allows us to address critical facets of accurate-segmenting lesions, namely, improving boundary delineation and addressing class imbalance. Training Optimisations like Automatic Mixed Precision (AMP) and Cyclic Learning Rate (CL) scheduling further refine the training process and reduce convergence problems and computational costs by 50%, improving efficiency.

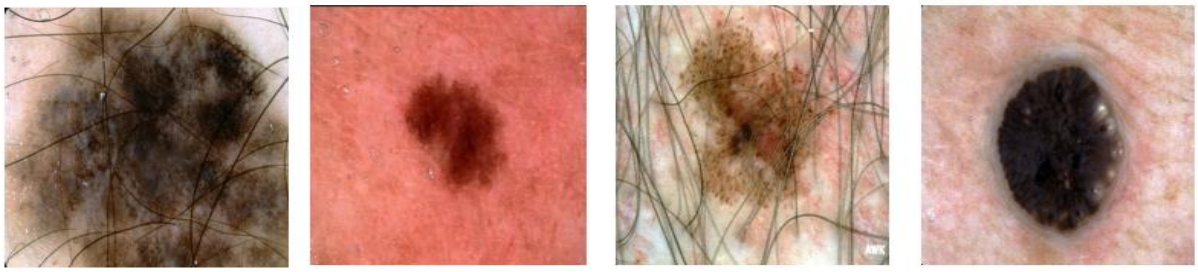
The main objective of this research is to develop a functional skin lesion segmentation system that achieves high accuracy and is clinically applicable, while also being implementable in healthcare settings with limited resources. Specifically, we: evaluate UNet++ using the EfficientNet-B5 encoder on the ISIC 2018 dataset, which includes 2,694 dermoscopic images with expert annotations; demonstrate performance gains over the baseline U-Net and O-Net models based on standard segmentation metrics; confirm the benefits of transfer learning and composite loss functions in improving boundary accuracy; and evaluate the practicality of this approach in the clinical domain by analyzing performance across different lesion varieties and imaging conditions.

This study represents a significant advance in creating a working prototype for AI skin diagnosis. It provides a viable solution to the shortcomings of the current state of the art, transformer-based methods, and addresses the challenges of computation and generalisation that preclude the integration of automated skin lesion segmentation systems into the clinical setting.

2. MATERIAL AND METHODS

2.1 Dataset

This analysis made use of the ISIC 2018 Challenge dataset, which contains a wide range of images of dermoscopic skin lesions. The labelling on the dataset is high quality. The dataset is designed to accommodate research spanning the automated classification of skin lesions and image segmentation. There are 2,694 images for training and 539 images for validation. Each image has a single high-quality mask for binary segmentation created by an experienced dermatologist. The mask outlines the lesion for each image in the dataset. The dataset is reflective of the variability which appears in real-world dermoscopic microscopy. Examples of the variability, and additional the additional lesions, include the following. The images present a wide range of heterogeneity in the lesions. Examples of heterogeneity in the lesions include differing shapes, differing levels of pigmentation, gap definitions on the borders, differing sizes, and differing clinical manifestations. A variety of standard clinical artefacts are included in the images: a cover with hair, markings with a ruler, a cover with air bubbles, shadows, and a variety of lighting and magnification. All of these artefacts also bear clinical relevance. The dataset provides an adequate standard for assessing the durability of a model. Figure 1(a) shows skin lesions obscured by hair, (b) with a tanned skin tone, (c) with bubbles and (d) large-sized lesions.



(a) (b) (c) (d)

Fig. 1: Representative samples from ISIC 2018 dataset showing diversity in lesion types, and imaging conditions including hair artifacts and varying skin tones [11].

2.2 Data augmentation

To enhance generalisation and reduce overfitting, a comprehensive data augmentation pipeline was implemented using the *Albumentations* library. The augmentation strategies were selected to simulate common dermoscopic variations and include: horizontal and vertical flips, translation of ±6.25%, scaling between 0.85 and 1.15 and random rotations of ±15° (each applied with 50% probability) for geometric transformation while photometric transformations include brightness and contrast adjustments (±15%), Gaussian blur, median blur and motion blur (maximum displacement of 3 pixels).

Coarse dropout introduces up to eight holes (size 32 × 32 pixels) to mimic hair occlusion, sensor noise, and other imaging disruptions. These augmentations significantly increased dataset variability while preserving lesion integrity, improving the model's robustness and boundary recognition.

2.3 Model architecture

This study employs an enhanced UNet++ architecture designed to improve multi-scale feature fusion and reduce the semantic gap between encoder and decoder representations.

The encoder is EfficientNet-B5, pre-trained on ImageNet. With its compound scaling strategy and optimized depth–width–resolution balance, EfficientNet-B5 provides high-quality hierarchical feature extraction, capturing subtle textural and morphological characteristics of dermoscopic lesions.

Nested skip connections, UNet++ integrates densely connected, nested skip pathways that progressively reduce the encoder–decoder feature disparity. This allows smoother feature propagation, enhanced gradient flow, and improved handling of fine lesion boundaries.

Spatial and Channel Squeeze-and-Excitation (SCSE) attention blocks were added to decoder stages to highlight clinically important features while suppressing irrelevant information. These attention mechanisms improve the network's focus on lesion structure, especially along irregular borders. Beyond segmentation, the encoder's global average-pooled features were fed into a lightweight classification head consisting of global average pooling, a fully connected dense layer, and sigmoid activation. This enabled simultaneous prediction of lesion malignancy (benign vs. malignant) alongside segmentation, thereby improving representation learning for both tasks.

To address class imbalance and optimise segmentation boundary precision, a weighted composite loss function was implemented:

- Dice Loss (50%): Encourages high overlap between predicted and ground truth segmentation masks.
- Binary Cross-Entropy (30%): Ensures accurate pixel-level classification.
- Focal Loss (20%): Forces the network to focus on difficult pixels, typically near lesion borders. This combination balances global region accuracy with fine-grained pixel-level discrimination.

During inference, multiple transformed versions of each image were generated for the test-time augmentation (TTA), such as horizontal, vertical, and diagonal flips at 90°, 180°, and 270° rotations. Prediction masks were obtained for each transformed version and averaged to produce a more stable final output and post-processing for final predictions underwent sigmoid activation, thresholding at 0.5, morphological closing to remove artifacts and resizing to the original image resolution.

2.4 Evaluation metrics

Segmentation performance was quantified using the dice similarity coefficient, Jaccard index (intersection over union), sensitivity, specificity, and accuracy. Classification was evaluated using sensitivity, specificity, and accuracy, F1-score and area under the ROC curve (AUC). All evaluations were conducted using the ISIC 2018 validation set.

3. RESULTS AND DISCUSSION

3.1 Classification and segmentation performances

The classification evaluates the model's ability to correctly categorise skin lesions as malignant or benign. Five key metrics were employed to comprehensively assess classification performance: sensitivity, specificity, accuracy, F1-score, and the AUC, which evaluates the model's discriminative capability across all classification thresholds. The proposed multi-task UNet++ architecture achieved sensitivity of 0.8900, specificity of 0.9200, accuracy of 0.9500, FI score 0.8870 and AUC 0.9278. O-Net had sensitivity 0.8825, specificity 0.9610, accuracy 0.9471, F1 score 0.8744, and AUC of 0.9260, and U-Net had sensitivity of 0.8512, specificity 0.9602, accuracy 0.9355, F1 score 0.8335 and AUC of 0.9102, respectively.

The segmentation task focuses on the precise delineation of lesion boundaries, which is essential for the accurate measurement of lesion size, shape, and asymmetry, key features in melanoma diagnosis. Two primary metrics were employed to evaluate segmentation quality: dice coefficient (DC) and the Jaccard Index: intersection over union (IoU). The proposed multi-task UNet++ architecture achieved sensitivity of 0.8900, specificity of 0.9200, accuracy of 0.9500, DC 0.8870 and IoU 0.8200, O-Net had sensitivity 0.8829, specificity 0.9610, accuracy 0.9471, DC 0.8704 and IoU of 0.8036, while U-Net had sensitivity of 0.8513, specificity 0.9572, accuracy 0.9355, DC 0.8314 and IoU with 0.7506, respectively, as seen in Tables 1 and 2

Table 1: Performance analysis for the classification of improved skin lesions

Model	Sensitivity	Specificity	Accuracy	F1-score	AUC
UNet++	0.8900	0.9200	0.9500	0.8870	0.9278
O-Net	0.8825	0.9610	0.9471	0.8744	0.9260
U-Net	0.8512	0.9602	0.9355	0.8335	0.9102
Model	Sensitivity	Specificity	Accuracy	F1-score	AUC
UNet++	0.8900	0.9200	0.9500	0.8870	0.9278

Table 2: Performance analysis for the segmentation of improved skin lesions

Model	**Sensitivity**	**Specificity**	**Accuracy**	**DC**	**IoU**
UNet++	0.8900	0.9200	0.9500	0.8870	**0.8200**
O-Net	0.8829	0.9610	0.9471	0.8704	**0.8036**
U-Net	0.8513	0.9572	0.9355	0.8314	**0.7506**
Model	Sensitivity	Specificity	Accuracy	DC	**IoU**
UNet++	**0.8900**	**0.9200**	**0.9500**	**0.8870**	**0.8200**

Each figure should have a caption. The caption should be concise and typed separately, not in the figure area. Figures should be self-explanatory. Information presented in the figure should not be repeated in the table. All symbols and abbreviations used in the illustrations should be defined clearly. Figure legends should be given below the figures.] A sample figure is given in Figure 1. The UNet++ model demonstrates significant improvements over U-Net and modest gains over O-Net in both classification and segmentation of skin lesions. In classification, it achieves 3.88% higher sensitivity, 5.35% higher F1-score, and 1.06% higher accuracy than U-Net, while slightly exceeding O-Net, reflecting a balanced detection of malignant lesions with minimised false positives and false negatives. In segmentation, UNet++ improves DC and IoU by 5.56% and 6.94% over U-Net, and 1.66% and 1.64% over O-Net, enabling precise delineation of lesion boundaries critical for morphological analysis. These improvements stem from nested skip connections that facilitate multi-scale feature aggregation, and the joint learning of classification and segmentation mirrors clinical workflows, highlighting UNet++ as a robust and reliable decision-support tool for melanoma detection.

3.1 Training convergence analysis

The convergence behaviour observed during training offers valuable insight into the learning dynamics and optimisation stability of the proposed UNet++ model. As shown in Figure 2, convergence curves are presented across training iterations (k) for three key performance metrics: accuracy/Matthews correlation coefficient (MCC) shown in blue, Dice coefficient in orange, and Intersection over Union (IoU) in green. All metrics exhibit rapid improvement in the early stages of training, followed by gradual stabilisation as the model approaches its optimal performance. In particular, the MCC displays a steady upward trend before plateauing at approximately 0.95, suggesting reliable learning behaviour and strong generalisation capability.

.

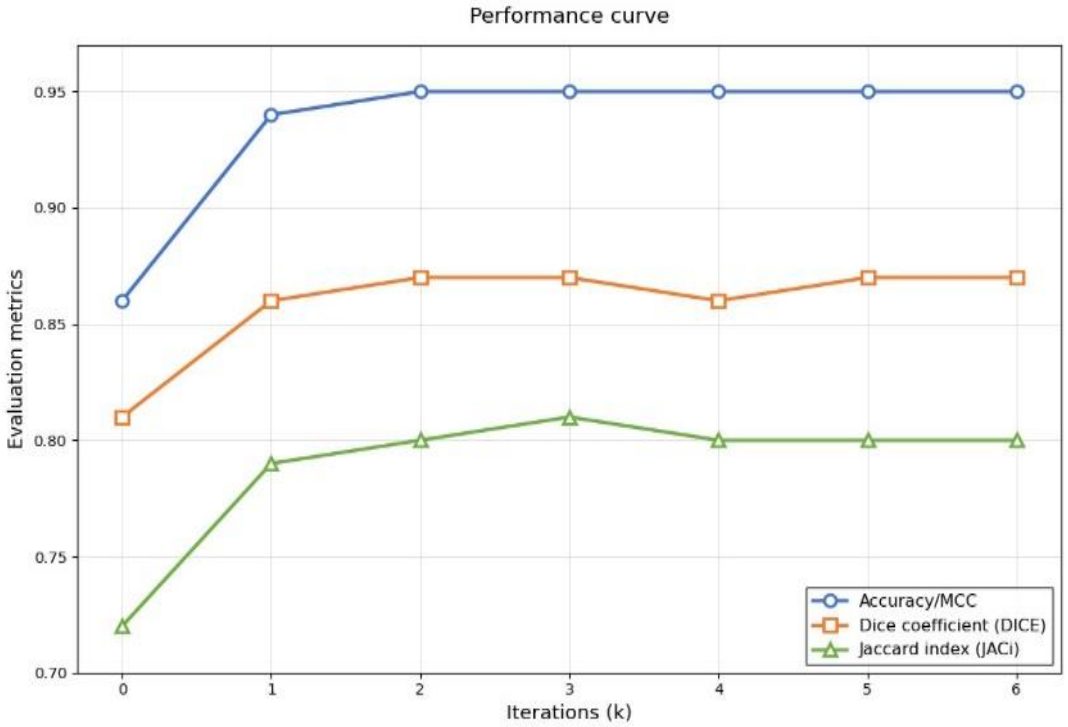


Fig. 2: Training convergence curves for UNet++ performance across iterations (k) MCC DC, and IoU

The DC curves also indicate a stable and well-behaved learning process, with rapid improvements during the early training stages followed by smooth convergence. No noticeable oscillations or divergence are observed. Importantly, signs of overfitting are absent, as reflected by the closely aligned progression of the MCC, Dice, and IoU curves throughout training. This consistent improvement across multiple evaluation metrics highlights the effectiveness of the composite loss strategy and supports the role of multi-task learning and attention-based feature refinement in enhancing segmentation performance**.**

3.3 Comparative analysis for the improved skin lesion segmentation

The proposed UNet++ model demonstrates meaningful improvements over both the classical U-Net and the more recent O-Net baselines. Although performance gains over O-Net are modest, the model achieves higher sensitivity (0.89 vs. 0.88) and improved segmentation metrics (DC: 0.89 vs. 0.87; IoU: 0.82 vs. 0.80), reflecting superior delineation of lesion boundaries and enhanced detection of subtle peripheral morphological changes relevant for early melanoma identification.

As shown in Figure 3, the comparative bar chart highlights the consistent performance advantage of the proposed model across sensitivity, accuracy, and overlap-based evaluation metrics, while achieving

specificity levels comparable to those of existing methods. This improvement can be attributed to several design choices, including the use of an EfficientNet-B5 encoder initialised with ImageNet pre-trained weights, the incorporation of SCSE attention mechanisms within the decoder, nested skip connections for enhanced feature fusion, and a composite loss function combining Dice, binary cross-entropy, and focal losses.

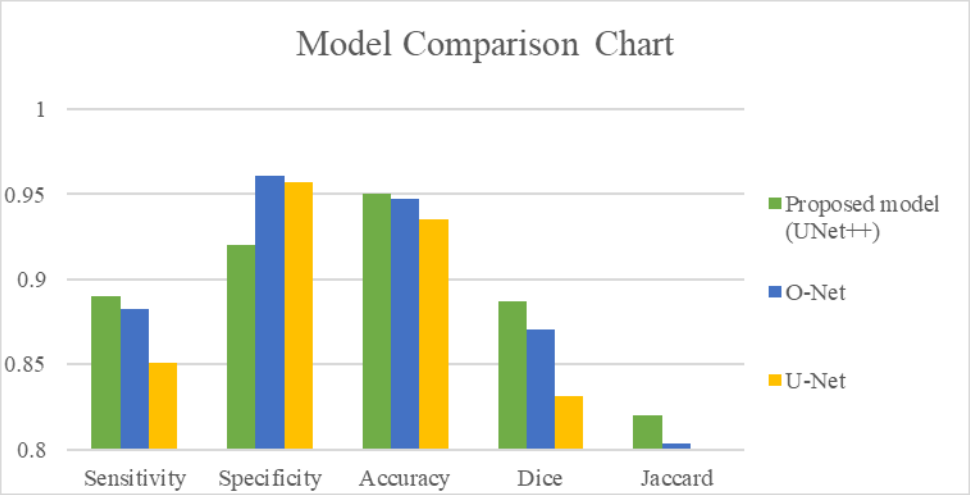


Fig. 3: Model comparison for the improved skin lesion segmentation

Although segmentation accuracy showed a slight decline when images contained severe artefacts, the predicted lesion masks remained clinically reasonable and preserved overall shape consistency. Performance reductions were mainly observed in cases involving very low-contrast lesions, images affected by dense or overlapping artifacts, and uncommon lesion types that are underrepresented in the ISIC 2018 dataset. Future studies should therefore focus on validating the UNet++ framework using larger, multi-centre datasets in order to improve model generalizability. Further performance gains may also be achieved by incorporating feature-enhancement mechanisms, such as attention-based modules or transformer architectures. Finally, evaluating the model within real-time clinical environments will be essential for determining its practical applicability and readiness for deployment in melanoma screening settings.

4. CONCLUSION

The refined UNet++ architecture combines EfficientNet-B5 encoding, multiscale feature aggregation, SCSE attention, and a composite loss function to optimise segmentation accuracy and classification fidelity. These changes facilitate boundary strengthening, domain and class imbalance mismatch reduction, and clinically robust pixel-wise and image-wise predictions. Unlike segmentation-only focused studies. Naqvi *et al.,* [10], this study investigates the segmentation and classification to determine the full diagnostic potential. As a result, comparable segmentation and classification outcomes to clinical benchmarks provide confidence in the accurate characterisation of lesions. Additionally, the proposed method, unlike the Transformer-based methods [10], Automatic Mixed Precision, to achieve notable accuracy while dramatically lowering the computational cost, proved highly adaptable to restrictive clinical environments.

ACKNOWLEDGEMENTS

We sincerely appreciate the support and cooperation of the management and staff of the Lagos State University. Their assistance and provision of a conducive environment greatly contributed to the successful completion of this study.

REFERENCES

1. Siegel, R. L., Miller, K. D., Wagle, N. S., and Jemal, A. (2023). Cancer statistics, 2023. CA Cancer J Clin, 73(1), 17-48. https://doi.org/10.3322/caac.21763.
2. Anderson, A. D. G., Carswell, S., Heath, H., Koutsis, J., and Guitera, P. (2024). Skin cancer referrals by nonmedical practitioners: a prospective observational study. Clinical and Experimental Dermatology, 49(9), 1048-1051. https://doi.org/10.1093/ced/llae115
3. Guy, G. P., Jr., Thomas, C. C., Thompson, T., Watson, M., Massetti, G. M., and Richardson, L. C. (2015). Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. MMWR Morb Mortal Wkly Rep, 64(21), 591-596.
4. Lee, Y.-S., Gu, H., Lee, Y.-H., Yang, M., Kim, H., Kwon, O., Kim, Y. H., and Kang, M.-Y. (2024). Occupational Risk Factors for Skin Cancer: A Comprehensive Review. Journal of Korean Medical Science, 39(42). https://doi.org/https://doi.org/10.3346/jkms.2024.39.e316.
5. Kittler, H., Marghoob, A. A., Argenziano, G., Carrera, C., Curiel-Lewandrowski, C., Hofmann-Wellenhof, R., Malvehy, J., Menzies, S., Puig, S., Rabinovitz, H., Stolz, W., Saida, T., Soyer, H. P., Siegel, E., Stoecker, W. V., Scope, A., Tanaka, M., Thomas, L., Tschandl, P. Halpern, A. (2016). Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy. J Am Acad Dermatol, 74(6), 1093-1106. https://doi.org/10.1016/j.jaad.2015.12.038.
6. Glazer, A. M., Rigel, D. S., Winkelmann, R. R., and Farberg, A. S. (2017). Clinical Diagnosis of Skin Cancer: Enhancing Inspection and Early Recognition. Dermatol Clin, 35(4), 409-416. https://doi.org/10.1016/j.det.2017.06.001
7. Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., Klode, J., Berking, C., Steeb, T., Enk, A. H., and von Kalle, C. (2018). Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. J Med Internet Res, 20(10), e11936. https://doi.org/10.2196/11936
8. Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., Gutman, D., Halpern, A., Helba, B., Hofmann-Wellenhof, R., Lallas, A., Lapins, J., Longo, C., Malvehy, J., Marchetti, M. A., Marghoob, A., Menzies, S., Oakley, A., Paoli, J. Kittler, H. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol, 20(7), 938-947. https://doi.org/10.1016/s1470-2045(19)30333-x
9. Chen, P., Huang, S., and Yue, Q. (2022). Skin Lesion Segmentation Using Recurrent Attentional Convolutional Networks. IEEE Access, 10, 94007-94018. https://doi.org/10.1109/ACCESS.2022.3204280
10. Naqvi, M., Gilani, S. Q., Syed, T., Marques, O., and Kim, H. C. (2023). Skin Cancer Detection Using Deep Learning-A Review. Diagnostics (Basel), 13(11). https://doi.org/10.3390/diagnostics13111911
11. Seixas, J. L., and Mantovani, R. G. (2016, 15-17 Dec. 2016). Decision Trees for the Detection of Skin Lesion Patterns in Lower Limbs Ulcers. 2016 International Conference on Computational Science and Computational Intelligence (CSCI),