JTE HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Enhancing Accuracy in Classification Models for Skin Disease Diagnosis Based on Segformer and ConvNeXt Approach

Hai-Duong Le[2], Van-Dung Hoang[1*]
[1]*Ho Chi Minh City University of Technology and Education, Vietnam*
[2]*MTI Technology Co., Ltd., Vietnam*

[*]Corresponding author. *Email: dunghv@hcmute.edu.vn*

ARTICLE INFO

Received:	16/01/2024
Revised:	23/02/2024
Accepted:	23/02/2024
Published:	28/08/2025

KEYWORDS

Skin Disease Diagnosis;
Machine Learning;
Segmentation;
Classification;
Medical Images.

ABSTRACT

This study introduces an innovative methodology to enhance the precision of skin disease diagnosis classification models by integrating segmentation results. Employing advanced machine learning techniques, our approach involves predicting lesion areas in skin images by combining SegFormer for skin lesion segmentation and backbone ConvNeXt for classifying skin images that consist of benign and malignant diseases. Based on training the SegFormer model for skin lesion segmentation, it achieved the IoU (intersection over union) ratio of 0.861 on the test set, outperforming the top 1 entry on the ISIC 2018 Leaderboards, which had an IoU of 0.802. Furthermore, our skin classification model uses image cropping to generate input images that emphasize damaged skin areas, eliminating redundant information. Leveraging the segmentation model's results, we define the bounding box for the lesion area, obtain a new image within the bounding box by adding padding, and then compare this new data with the original data. The disease classification model, using ConvNeXt as its backbone, exhibited superior performance on the new dataset compared with the original dataset, achieving a higher accuracy of 1.61%, precision of 26.42%, and recall of 26.49%. This research paves the way for novel approaches to address disease diagnosis challenges in medical images, particularly in skin diseases. It can improve the performance of classification models when trained on image datasets that do not have synchronization during acquisition.

Doi: https://doi.org/10.54644/jte.2025.1522

1. Introduction

Skin diseases pose a significant global health challenge, with skin cancer being the most widespread cancer, not only in the United States and Australia but also in the world. Recent statistics emphasize the seriousness of this issue. Lim et al.'s study [1] revealed that 84.5 million Americans, constituting one in four, struggle with various skin diseases, resulting in substantial healthcare costs estimated at $75 billion. These costs include medical treatments, preventive measures, and the use of both prescription and over-the-counter drugs.

The impact of skin cancer is particularly distressing, as highlighted by The Skin Cancer Foundation [2] The foundation reports a grim reality, with two individuals succumbing to skin cancer every hour in the United States. Additionally, one in five Americans is projected to develop skin cancer by the age of 70. In Australia, Spotscreen [3] reveals an even more alarming scenario, with two out of three individuals diagnosed with skin cancer by the age of 70. Shockingly, over 2,000 people receive treatment for skin cancer every day, amounting to approximately 750,000 cases annually. Despite these daunting statistics, a glimmer of hope emerges when detected early, which can support the 5-year survival rate for melanoma is 99% [1]. This underscores the crucial importance of developing and implementing advanced, automated diagnostic tools to enable early detection and improve outcomes in the battle against skin diseases, particularly skin cancer.

However, this transition faces a critical obstacle – the scarcity and inconsistency of data, especially concerning skin cancer. In this paper, a new method is proposed to effectively address the challenges associated with imbalanced data and improve the classification performance of machine learning models. This approach combines SegFormer for skin lesion segmentation and ConvNeXt to classify skin images into benign and malignant diseases. Training the SegFormer model for skin lesion segmentation achieved an IoU ratio of 0.861 on the test set, surpassing the top entry on the ISIC 2018 Leaderboards. Leveraging the segmentation model's results, we define bounding boxes for the lesion areas by identifying the contours of the lesions. We then crop the original images using these bounding boxes to create refined datasets for training. The ConvNeXt classification model demonstrates excellent performance on these datasets, showing higher accuracy, precision, and recall than the original dataset. Our research contributes to the field of automated diagnostics and provides an approach for medical image classification problems, particularly in the context of skin diseases.

2. Related work

Automating skin disease identification and classification not only expedites diagnostic processes but also helps reduce costs and increase accessibility to patients. Traditional techniques for skin disease image classification, as explored by Amarathunga et al. [4], involve a time-consuming feature selection process that demands meticulous attention to relevant features. Their expert system, which focused on three diseases, used a two-unit structure for data and image processing, showcasing promising results with the MLP classifier. Chakraborty et.al.[5] proposed a hybrid model that integrates NSGA-II and ANN for skin lesion diagnosis, employing the bag-of-features approach with SIFT and clustering algorithms. Chatterjee et al. [6] adopted spatial and frequency domain techniques using cross-correlation and SVM with the RBF kernel for the precise classification of skin images on malignant and benign lesions.

In contrast, approaches based on deep learning for image classification of skin disease, particularly CNNs, revolutionize this aspect by autonomously learning features efficiently. CNNs intelligently select filters during feature extraction, presenting a marked improvement over the manual selection method. In the domain of deep learning, Esteva et al. [7] demonstrated that CNNs, specifically the InceptionV3 architecture, could rival dermatologists in identifying malignant lesions. Subsequent studies by Zhang et al. [8] and Sun et al. [9] further explored InceptionV3 and other CNN architectures, emphasizing the significance of training data and cautioning against misclassification due to multiple lesions. Gessert et al. [10] presented a method based on patches using CNN architectures such as Inception, SE-Resnext50, and DenseNet for fine-grain disease classification using high-resolution image patches. Rehman et al. [11] presented a CNN architecture using 16 filters for malignant and benign skin disease classification. Kulhalli et al. [12] addressed class imbalance using a hierarchical approach with image augmentation, suggesting further refinement and ensemble-based methods for enhanced performance.

Additionally, some pre-trained models, particularly InceptionV3 [13], ResNet [14], VGG16 [15], AlexNet [16], and so on have been investigated. These models, trained on extensive datasets comprising millions of diverse images, exhibit adaptability in skin disease diagnosis through transfer learning or fine-tuning. The automatic feature extraction capabilities of CNNs, coupled with the use of pre-trained models, contribute to the effectiveness of image classification for skin disease diagnosis systems, which are potentially for enhanced diagnostic accuracy and efficiency in dermatological applications.

Advanced methods using both segmentation and classification have been explored to improve medical image classification. Studies integrating segmentation outputs to classify diseases, such as lung diseases through CT images [17], [27], heart disease via cine MRI [18], or brain damage via blood vessel occlusion [19], exemplify the integration of these methods. Some skin disease classification models incorporate segmentation as a pre-processing step, focusing on damaged skin areas or creating regions of interest (ROIs) [20]-[21]. These studies showcase the evolving landscape of research through strategic combinations of segmentation and classification methodologies.

Despite the promising strides in the integration of segmentation and classification, several challenges remain in the field of skin disease diagnosis using deep learning. The availability of labeled skin disease data has some limitations and imbalanced datasets. Deep learning algorithms, which are keen on data from diverse sources, face difficulties when applied to images captured under different conditions, with different shooting angles, and especially when the rate of lesion on the image is different. Our research

JTE
HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

proposes an approach for effectively addressing the challenges associated with inconsistent data and improving the classification performance of skin diseases.

3. Proposed Approach

Our study presents a new approach methodology, which is aimed at enhancing the efficacy of skin lesion classification, with potential applicability to various medical image classification challenges. The proposed approach comprises three stages. In the first stage, a segmentation model is employed to accurately delineate and identify lesion areas within medical images. The second stage involves generating a region of interest for the lesion by cropping the original image using the bounding box surrounding the affected skin, along with an additional padding space. Finally, in the third stage, new images are incorporated into the image classification model. The general flowchart of systems is illustrated in Figure 1.

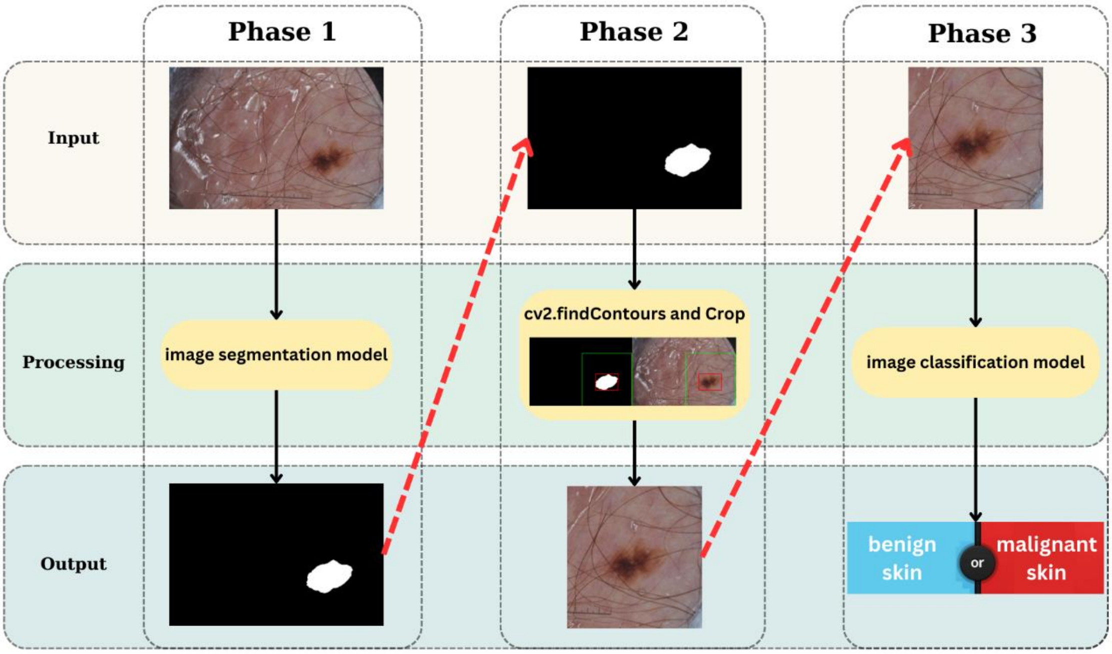

Figure 1. *Overview approach for skin lesion analysis which segmentation, contour extraction, cropped lesion images, and classification.*

In the initial phase of skin lesion analysis, the Segformer [22] architecture was employed for lesion segmentation. The primary objective is to accurately delineate the boundaries of the skin lesions. The methodology uses Segformer for semantic segmentation, where pixels are classified into background and foreground classes, creating a spatial segmentation map. This step is significant in identifying and constraining the extent of the damaged skin area within the entire image.

Subsequently, a comprehensive view of the skin lesions, including the perilesional area, was generated using cropped lesion image generation with contour padding. This involves cropping the lesion image to establish a region of interest (ROI). The identification of contours is achieved using cv2.findContours. To capture the surrounding skin context effectively, the bounding box is padded by 20% of the image size. Figure 2 shows how the ROI region of interest is formed. Starting from the original image, the segmentation model is applied to obtain the lesion mask. Then the red bounding box is created by using cv2.findContours to identify the extent of the damaged area in the image. Next, The blue bounding box is created by adding padding to the red bounding box. Finally, the image area corresponding to the blue bounding box in the original image is cropped to generate a new image with more focused content on the damaged skin area.

For the final phase of our methodology, we introduce a model termed SkinClassifier, which leverages the ConvNeXt [23] backbone to extract relevant information. Subsequently, the extracted information is passed through two fully connected layers to generate the final prediction. The model is trained on the output derived from phase 2, ensuring optimal consistency in the trained images by emphasizing key regions within the lesion images and preserving the scale between the skin lesions and normal skin. This approach is especially beneficial when the difference between lesions and normal skin varies across images in the dataset. The combination of these three procedural steps establishes a comprehensive framework for the study of skin lesions, encompassing segmentation, contour extraction, cropped image generation, and classification.

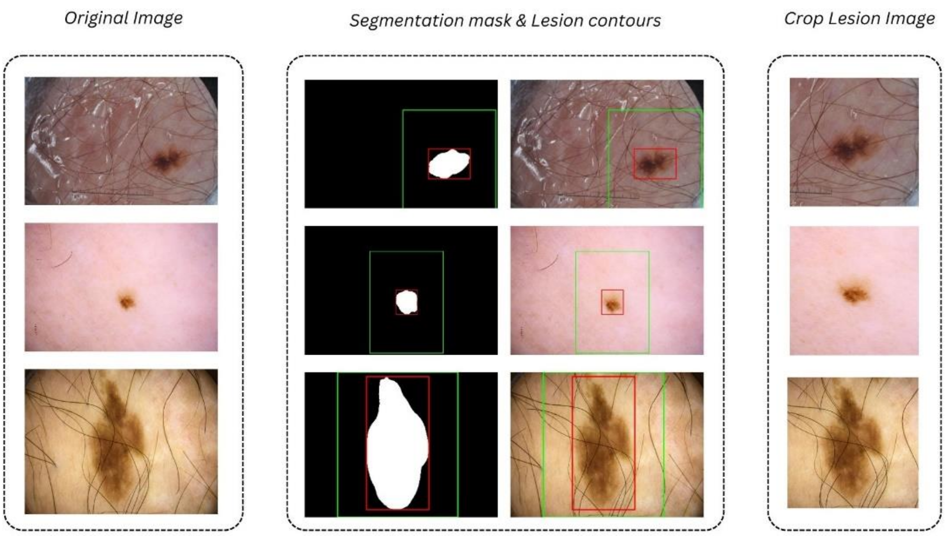

Figure 2. *Some examples of original skin images, segmentation, and ROI of the second phase.*

4. Results and Discussion

We conducted experiments to evaluate the performance of the Segformer segmentation model on the ISIC2018 dataset [24], which consists of three sets for training, validation, and testing, comprising 2,594, 100, and 1,000 images, respectively. Each skin image is accompanied by a corresponding grayscale image representing the mask of the damaged area. Subsequently, we applied the Segformer model to implement our proposed method on the ISIC 2020 dataset [25]. The SkinClassifier image classification model is trained on two datasets: the original dataset and the dataset augmented by the proposed method. The comparative results between these two datasets illustrate the performance of our integrated method.

The ISIC 2020 dataset [25], comprises a total of 33,126 images, with 32,542 images depicting benign skin and 584 images representing malignant skin. We partitioned the dataset into training and testing sets with an 80/20 ratio. The training set included 26,033 images of benign skin and 467 images of malignant skin, whereas the testing set included 6,509 images of benign skin and 117 images of malignant skin. To address the imbalanced nature of the dataset, we employed image transformation methods such as rotation, flipping, blurring, and noise addition techniques to generate new malignant skin images, thereby mitigating the class imbalance in the training set and ensuring equal representation between benign and malignant skin groups. Figure 3 depicts an example of new malignant skin images generated from an original image.

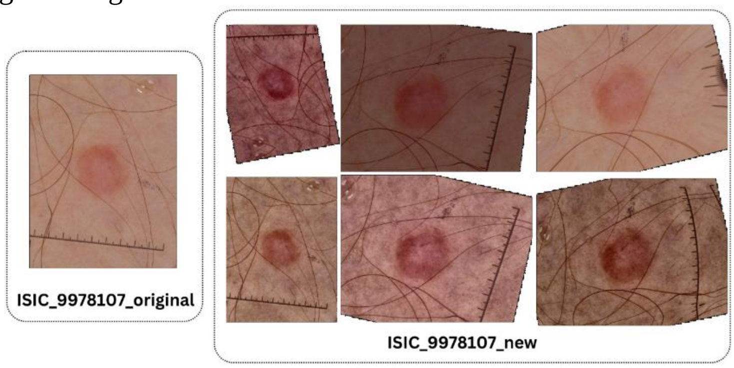

Figure 3. *Results of newly generated malignant skin images.*

To evaluate the effectiveness of the proposed method, we compared the performance of the classification model when training on the original dataset, which we call "Original Image" which are images taken from the ISIC 2020 set. The second dataset is "Crop Lesion Image" which includes images from the ISIC 2020 set cropped to obtain ROI using the proposed method. The performance of the Segformer model trained on the ISIC 2018 set will affect the quality of the "Crop Lesion Image" set. The images on the two datasets are divided and processed in the same pipeline, and the performance difference between the two Skin Classifier training models on the two datasets will demonstrate the effectiveness of our method.

For the evaluation metric, we employed commonly used performance measures to assess the classification and segmentation results. For the classification task, we used metrics such as accuracy, precision, recall score to evaluate the model efficiency to distinguish between benign and malignant skin lesions. In the context of skin lesion segmentation, we measured the model's performance using metrics such as the IoU and the dice coefficient. These metrics provided a quantitative assessment of the segmentation accuracy and delineation of lesion boundaries.

The image segmentation model, employing the Segformer architecture, exhibited exceptional performance on the ISIC2018 dataset [24], achieving an IoU of 0.861 and a Dice coefficient of 0.925 on the test set. Notably, these results surpassed the top-performing model on the ISIC 2018 Leaderboards [26], named MaskRcnn2+segmentation, which held an IoU index of 0.802. Table 1 presents the ranking of the top 5 methods with the highest IoU on the test set at the time the SIC 2018 challenge permitted submissions. All these methods plateaued at an IoU of approximately 0.8, while our Segformer model achieved an IoU of 0.861, showcasing its superiority over previous models. It is essential to emphasize that the models in this ranking were developed during the ISIC 2018 challenge timeframe, whereas Segformer was introduced in 2021. The superior performance of the Segformer model compared to the methods on the ISIC 2018 Leaderboards is attributed to advancements in its architecture.

Table 1. *ISIC 2018 Leaderboards - Task 1: Lesion Boundary Segmentation [26]*

Rank	Team	Approach Name	IoU
1	MT	MaskRcnn2+segmentation	0.802
2	Holidayburned	ensemble_with_CRF_v3	0.799
3	imsight	Automatic Skin Lesion Segmentation by DCNN	0.799
4	Tencent Youtu Lab	Skin Lesion Segmentation with Adversarial Learning	0.798
5	NMN_team	segmentation_ensembleALL_Th0.80_Tl0.65	0.796

Building upon the accurate segmentation, the SkinClassifier model, with a ConvNeXt backbone, capitalized on the output of the Segformer model to crop the central image area. This innovative approach resulted in remarkable classification performance on the ISIC dataset [25], attaining an accuracy of 98.21%, precision of 49.68%, and recall of 66.66%. This performance represents a significant leap beyond traditional methods, showing higher Accuracy by 1.61%, higher Precision by 26.42%, and higher Recall by 26.49%. In the context of this study, 'traditional methods' refers to the standard practice of training and evaluating models on raw and uncropped images. The detailed results obtained after 10 epochs on the two datasets are presented in Table 2. These results collectively demonstrate the effectiveness of our proposed methodology, underscoring its potential for advancing the field of skin disease diagnosis through improved segmentation and classification models.

Table 2. *Results of the SkinClassifier model on the test set of two datasets.*

Dataset	Accuracy	Precision	Recall
Original Image	96.60%	23.26%	40.17%
Crop Lesion Image (proposed method)	**98.21%**	**49.68%**	**66.66%**

5. Conclusions

Our study presents a method to enhance the performance of skin disease classification models by leveraging the outcomes of image segmentation models. Experimental results substantiate that the proposed method effectively addresses the challenge of asynchrony in collecting skin disease data and can be extended to other medical image classification problems. Notably, our findings underscore the efficacy of the Segformer architecture, surpassing existing benchmarks in skin lesion segmentation, particularly outperforming the top-ranking model on the ISIC Leaderboard 2018. The SkinClassifier model, employing the ConvNeXt backbone and using output from Segformer, exhibits better classification performance on the ISIC dataset, demonstrating higher accuracy, precision, and recall compared with traditional methods. While our work contributes to advancements in computer vision for the medical field, future research should prioritize several key areas. It is crucial to expand our method's training and evaluation on diverse datasets to enhance generalizability across various skin conditions. Participation with industry experts and healthcare professionals will refine performance models for real-world deployment and address practical challenges. In addition, improving model interpretability is vital for gaining broader acceptance. By integrating explainability features and ongoing refinement efforts, we aim to revolutionize skin disease diagnosis and medical image evaluation, paving the way for impactful applications in dermatology and beyond.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

[1] H. W. Lim *et al.*, "The burden of skin disease in the United States," *Journal of the American Academy of Dermatology*, vol. 76, no. 5, pp. 958–972, 2017.

[2] The Skin Cancer Foundation, "Skin Cancer Facts," Since 1979, The Skin Cancer Foundation has set the standard for educating the public and the medical community about skin cancer. [Online]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

[3] Spotscreen, "Skin Cancer Facts," As award-winning industry leaders, Spotscreen specializes in best-practice corporate skin cancer screening programs across Australia. [Online]. Available: https://www.spotscreen.com.au/info-centre/skin-cancer-information/skin-cancer-facts/

[4] A. Amarathunga, E. Ellawala, G. Abeysekara, and C. Amalraj, "Expert system for diagnosis of skin diseases," *International Journal of Scientific & Technology Research*, vol. 4, no. 01, pp. 174–178, 2015.

[5] S. Chakraborty *et al*., "Image based skin disease detection using hybrid neural network coupled bag-of-features," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, 2017, pp. 242–246.

[6] S. Chatterjee, D. Dey, S. Munshi, and S. Gorai, "Extraction of features from cross correlation in space and frequency domains for classification of skin lesions," *Biomedical Signal Processing and Control*, vol. 53, p. 101581, 2019.

[7] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[8] X. Zhang, S. Wang, J. Liu, and C. Tao, "Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge," *BMC Medical Informatics and Decision Making*, vol. 18, no. 2, pp. 69–76, 2018.

[9] X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI, vol. 14, pp. 206–222, Springer, 2016.

[10] N. Gessert *et al.*, "Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 495–503, 2019.

[11] M. Rehman, S. H. Khan, S. D. Rizvi, Z. Abbas, and A. Zafar, "Classification of skin lesion by interference of segmentation and convolution neural network," in *2018 2nd International Conference on Engineering Innovation (ICEI)*, IEEE, 2018, pp. 81–85.

[12] R. Kulhalli, C. Savadikar, and B. Garware, "A hierarchical approach to skin lesion classification," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, 2019, pp. 245–250.

[13] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 2017, pp. 783–787.

[14] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," arXiv preprint arXiv:1603.08029, 2016.

[15] D. Theckedath and R. Sedamkar, "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks," *SN Computer Science*, vol. 1, pp. 1–7, 2020.

[16] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size," 2016, doi: 10.48550/arXiv.1602.07360.

[17] P. Lobo and S. Guruprasad, "Classification and segmentation techniques for detection of lung cancer from CT images," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, 2018, pp. 1014–1019.

[18] A. Ammar, O. Bouattane, and M. Youssfi, "Automatic cardiac cine MRI segmentation and heart disease classification," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101864, 2021.

[19] B. Shahangian and H. Pourghassem, "Automatic brain hemorrhage segmentation and classification in CT scan images," in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, IEEE, 2013, pp. 467–471.

[20] V. D. Hoang, X. T. Vo, K. A. Phu, and K. H. Jo, "Fusion of segmentation and classification for improving skin disease diagnosis," in *International Conference on Green Technology and Sustainable Development*, Springer, 2022, pp. 144–154.

[21] R. Sumithra, M. Suhil, and D. Guru, "Segmentation and classification of skin lesions for disease diagnosis," *Procedia Computer Science*, vol. 45, pp. 76–85, 2015.

[22] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.

[23] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 11966-11976, doi: 10.1109/CVPR52688.2022.01167.

[24] N. Codella *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," 2019, doi: 10.48550/arXiv.1902.03368.

[25] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific data*, vol. 8, no. 1, p. 34, 2021.

[26] International Skin Imaging Collaboration (ISIC), "ISIC 2018 Leaderboards," ISIC Challenge. [Online]. Available: https://challenge.isic-archive.com/leaderboards/2018/

[27] N. N. Tran *et al.*, "Segmentation on chest CT imaging in COVID-19 based on the improvement attention U-Net model," in *New Trends in Intelligent Software Methodologies, Tools and Techniques*, IOS Press, 2022, pp. 596-606.

Hai-Duong Le received bachelor's degree in Information Technology from Ho Chi Minh City University of Technology and Education, Vietnam, in 2024. He is working at MTI Technology Co., Ltd.
Email: lehaiduongbmt@gmail.com. ORCID: https://orcid.org/0009-0000-7534-9862

Van-Dung Hoang received a Ph.D. degree in Electrical and Computer Engineering from University of Ulsan, Ulsan city, Korea, in 2015. After a year of experience with Telecom SudParis as a Postdoctoral Research Fellow, he joined the Faculty of Information Technology, HCMC University of Technology and Education, Ho Chi Minh City, Vietnam, where he is currently serving as a professor and head of Artificial Intelligence Department. He has a total academic teaching experience of 18 years with more than 85 publications in reputed international conferences, journals, and book chapter contributions (Indexed By: ISI, WoS, Scopus, ACM DL, DBLP). His research interests cover a wide area, which focuses on computer vision, robotics, medical image processing, autonomous vehicles, and ambient intelligence.
Email: dunghv@hcmute.edu.vn. ORCID: https://orcid.org/0000-0001-7554-1707