A Lightweight Model to Skin Disease Recognition

To Huu Nguyen[1], Ma T. Hong Thu[2], Do Thanh Mai[3], Trang Phung T. Thu[3], An Dang[4], Duc-Quang Vu[5]
[1] Thai Nguyen University of Information and Communication Technology, Thai Nguyen, Vietnam
[2] Tan Trao University, Tuyen Quang, Vietnam
[3] Thai Nguyen University, Thai Nguyen, Vietnam
[4] Faculty of Computer Science, Phenikaa University
[5] Thai Nguyen University of Education, Thai Nguyen, Vietnam
Correspondence: Duc-Quang Vu, quangvd@tnue.edu.vn
Received: 14/2/2024, revised: 16/5/2024, accepted: 27/5/2024
Digital Object Identifier: 10.32913/mic-ict-research-vn.v2024.n1.1265

Abstract: Skin disease has become increasingly prevalent, emerging as one of the most widespread conditions. It significantly affects human health and even causes skin cancer and death. Therefore, there are many methods have been proposed to solve this issue recently, especially deep learning-based methods. However, these state-of-the-art methods seem to only focus on how to achieve better performance and ignore the issue of inference time. Specifically, deep learning-based methods usually build very deep with a huge of model size and computational cost. As a result, this becomes very difficult to deploy these models on devices with no GPU support. In this study, we introduce a proficient and lightweight model designed to address this issue, leveraging the Mobilenet architecture. Our experimental findings demonstrate that the suggested network delivers comparable performance to contemporary cutting-edge techniques across diverse benchmark datasets, including HAM10000, International Skin Imaging Collaboration 2017, and International Skin Imaging Collaboration 2019. Notably, our approach utilizes merely 0.2 million parameters and 0.3 GFLOPs for image classification. This attribute holds substantial importance for deploying the model on edge devices lacking GPU support.

Keywords: *skin disease, skin lesion classification, lightweight network, MobileNet*

Tiêu đề: **Một mô hình nhẹ để nhận biết bệnh về da**

Tóm tắt: Bệnh về da ngày càng trở nên phổ biến, nổi lên như một trong những tình trạng phổ biến nhất. Nó ảnh hưởng đáng kể đến sức khỏe con người, thậm chí gây ung thư da và tử vong. Vì vậy, gần đây có rất nhiều phương pháp được đề xuất để giải quyết vấn đề này, đặc biệt là phương pháp dựa trên học sâu. Tuy nhiên, những phương pháp tiên tiến này dường như chỉ tập trung vào việc làm thế nào để đạt được hiệu suất tốt hơn mà bỏ qua vấn đề về thời gian suy luận. Cụ thể, các phương pháp dựa trên học sâu thường xây dựng rất sâu với kích thước mô hình và chi phí tính toán rất lớn. Do đó, việc triển khai các mô hình này trên các thiết bị không hỗ trợ GPU trở nên rất khó khăn. Trong nghiên cứu này, chúng tôi giới thiệu một mô hình gọn nhẹ và hiệu quả được thiết kế để giải quyết vấn đề này tận dụng kiến trúc Mobilenet. Các kết quả thử nghiệm của chúng tôi chứng minh rằng mạng được đề xuất mang lại hiệu suất tương đương với các kỹ thuật tiên tiến hiện đại trên các bộ dữ liệu chuẩn khác nhau, bao gồm HAM10000, International Skin Imaging Collaboration 2017 và International Skin Imaging Collaboration 2019. Đáng chú ý hơn, phương pháp của chúng tôi chỉ sử dụng 0,2 triệu thông số và 0,3 GFlops để phân loại hình ảnh. Điều này có tầm quan trọng đáng kể trong việc triển khai mô hình trên các thiết bị cạnh không có GPU hỗ trợ.

Từ khóa: *bệnh ngoài da, phân loại tổn thương da, mạng nhẹ, MobileNet*

I. INTRODUCTION

Skin cancer stands as the most prevalent among all cancer types, surpassing the cumulative total of other cancer categories combined [1]. The occurrence of skin cancer is particularly prominent among individuals of Caucasian descent, with an incidence rate of approximately 200 cases per 100,000 population. In Vietnam, although the prevalence of skin cancer is comparatively lower in comparison to Caucasians, there has been a noticeable rise in cases. This trend is particularly notable among men, with an incidence of about 3.2 cases per 100,000 population, and among women, with a rate of 3.1 cases per 100,000 individuals. The most lethal variant of skin

cancer, melanoma, is projected to escalate to nearly half a million cases by 2040, marking a 62 percent surge from 2018 [2]. Astonishingly, skin cancer claims a life every four minutes, prompting dermatologists to classify the surge in skin cancer occurrences as a worldwide pandemic.

Prompt intervention, particularly for melanoma, is pivotal for achieving favorable survival outcomes amidst the burgeoning instances [3]. The chief discernible trigger for skin cancer is prolonged exposure to ultraviolet radiation. Dwindling ozone levels contribute to escalated high-level ultraviolet radiation, consequently heightening the risk associated with natural sunlight exposure.

Dermatoscopy stands as a widely employed imaging methodology, utilizing immersion fluid to amplify light and enable visualization of the skin's surface [4]. Nonetheless, the precision of its diagnostic outcomes hinges greatly on the experience of the dermatologist [5, 6]. The dearth of specialized medical personnel, particularly in developing or economically disadvantaged nations, can profoundly impede the timely management of skin cancer cases. Consequently, the necessity for automated, remote diagnostic solutions is assuming heightened significance. The scarcity of specialized medical personnel, especially in developing or economically disadvantaged nations, can significantly hinder the prompt management of skin cancer cases. Consequently, the classification of skin lesions has emerged as a burgeoning realm of study, spurred by the increasing integration of deep learning techniques in the realm of medical image analysis.

The advent of deep learning models has brought about a transformative shift in conventional machine learning applications, offering unparalleled performance enhancements and the capacity to autonomously learn intricate information. Deep neural networks have proven successful across diverse domains, spanning image classification [7, 8], object detection [9], action recognition [10, 11], speech recognition [12, 13], and video prediction [14, 15], among others. There are various methods proposed in skin lesion classification (SLC) such as Zhang et al. [16], Datta et al. [17], TALT [18] and Huynh et al [19]. However, many of these methodologies lean towards utilizing exceedingly large models to attain peak performance, often neglecting secondary factors such as memory usage and runtime efficiency. For example, the ResNet-50 uses 45M parameters and 8GFLOPs, VGGNet uses 133M parameters and 8GFLOPs, etc. This seriously affects the process of deploying these models on embedded/mobile devices without GPU support. In this study, we introduce an efficacious lightweight network that prioritizes minimized memory and computational expenses. This aspect bears substantial significance, particularly when deploying the model on embedded devices devoid of GPU support. To be precise, we present an innovative framework rooted in MobileNet architecture [20] aimed at curbing the model's size and computational overhead in the context of SLC. The principal contributions of this study can be succinctly outlined as follows:

1) Our proposition involves the introduction of a streamlined and resource-efficient network for SLC, built upon the MobileNet architecture.
2) Empirical findings from our experiments conclusively demonstrate the remarkable reduction in both model size and computational expenditure achieved by the suggested network, standing in stark contrast to prevailing deep models at the forefront of the field.

II. PROPOSED MODEL

Table I
THE PROPOSED LIGHTWEIGHT NETWORK FOR SKIN LESION CLASSIFICATION BASED ON MOBILENET.

Type/Stride	**Kernel size × No. of filters**	**Output size**
Input layer		(224,224,3)
Conv2D / s=(2,2)	(3,3) × 8	(112,112,8)
Depthwise / s=(1,1)	(3,3)	(112,112,8)
Pointwise / s=(1,1)	(1,1) × 16	(112,112,16)
Depthwise / s=(2,2)	(3,3)	(56,56,16)
Pointwise / s=(1,1)	(1,1) × 32	(56,56,32)
Depthwise / s=(1,1)	(3,3)	(56,56,32)
Pointwise / s=(1,1)	(1,1) × 32	(56,56,32)
Depthwise / s=(2,2)	(3,3)	(28,28,32)
Pointwise / s=(1,1)	(1,1) × 64	(28,28,64)
Depthwise / s=(1,1)	(3,3)	(28,28,64)
Pointwise / s=(1,1)	(1,1) × 64	(28,28,64)
Depthwise / s=(2,2)	(3,3)	(14,14,64)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(1,1)	(3,3)	(14,14,128)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(1,1)	(3,3)	(14,14,128)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(1,1)	(3,3)	(14,14,128)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(1,1)	(3,3)	(14,14,128)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(1,1)	(3,3)	(14,14,128)
Pointwise / s=(1,1)	(1,1)×128	(14,14,128)
Depthwise / s=(2,2)	(3,3)	(7,7,128)
Pointwise / s=(1,1)	(1,1)×256	(7,7,256)
Global AVE Pooling		(256,)
Output layer	K	(K,)

In this part, we first provide a concise overview of the problem setup. Subsequently, we introduce our proposed

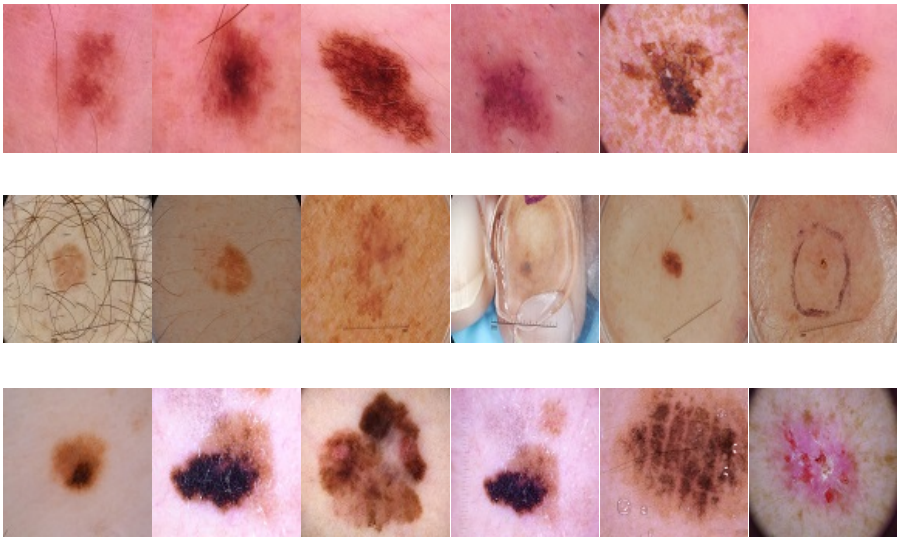

Figure 1. Several example images on three datasets. In which, the top of the figure is the samples in HAM10000, the middle is the samples in ISIC 2017 and the bottom is the samples in ISIC 2019.

lightweight network designed for skin disease recognition, utilizing the MobileNet backbone as its foundation.

1. Problem Setup

Given a dataset denoted by $\mathcal{D} = \{(x_1, y_1), ..., (x_N, y_N)\}$. In which, x_i denotes the RGB training image and y_i is its corresponding label with $i = 1, 2, ...N$ and N is the total of samples in $\mathcal{D}$. Note that $x_i \in \mathbb{R}^{H,W,C}$ where H, W, C as the height, width, and channel dimension of the image, respectively, and $y_i \in \{1, 2, ..., K\}$ where K is the total of classes. We denote $p_i = F_\theta(x_i)$ where F_θ is the proposed network and p_i is the distribution probability vector of the network F_θ, θ indicates the trainable set of parameters. We aim to minimize the distance between the output prediction $\hat{y}_i$ from F_θ and the ground-truth label y_i i.e., $\hat{y}_i = \arg\max(p_i) \approx y_i$. To do that, we utilize the standard cross-entropy (CE) to perform this task as follows:

$$CE(y_i, \hat{y}_i) = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{1}$$

2. Lightweight Network

Conventional Convolution. is computed over the entire depth (channel). Therefore, the total parameters of the network increase significantly depending on the depth of the previous layer. For example, we assume that the input of a standard convolution is $h \times w \times c$ we need a set of $k \times k \times c$ parameters to perform the convolution operator and generate an output size of $h' \times w' \times 1$. Therefore, we need c' different filters to create the output size of $h' \times w' \times c'$. From that, the total parameters used for a normal convolution will be $c' \times k \times k \times c$

The number of parameters is negligible if the depth of the input channel is small (usually in the first layers). However, when the depth increases gradually toward the final layers of the Convolutional Neural Network (CNN), the number of parameters of the model will increase significantly. This increase in parameters creates cumbersome models that take up a huge of memory and affect computation speed.

Depthwise Convolution. Inspired by MobileNet, we adopt Depthwise Convolution to reduce the computational cost by performing convolution on each slice. Specifically, each channel will apply a different filter and not share parameters at all. To produce a pixel point on the output, regular convolution uses $k \times k \times c$ operations meanwhile Depthwise Convolution only utilizes $k \times k$.

Pointwise Convolution. is adopted to change the depth channel of the output from c to c' by using 1 Convolution. The detail of our proposed framework is described in Table I. In which, s denotes the stride of the convolution operator, Global AVE Pooling is the Global Average Pooling layer, Depthwise is the Depthwise convolution operator and Pointwise is the pointwise convolution operator.

III. EXPERIMENT

In this section, we commence by introducing several standard datasets commonly employed in skin disease recognition. Following that, we delve into a detailed explanation of the implementation of our proposed network. Concluding this part, we provide the results of various experiments,

Table II
THE COMPARISON BETWEEN OUR PROPOSED LIGHTWEIGHT NETWORK AND RECENT MODELS ON THE HAM10000 DATASET INCORPORATES THE ACCURACY VALUES ACROSS SEVEN DISTINCT CLASSES.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC
DenseNet [21]	0.975	0.993	0.960	0.851	0.963	0.975	0.993
VGGNet [22]	0.949	0.977	0.93	0.847	0.925	0.954	0.972
ResNet [7]	0.980	0.997	0.948	0.973	**0.961**	0.974	0.995
Inception-v4 [23]	**0.993**	0.997	0.970	0.973	0.965	0.984	**1.000**
Inception-v4 [23] + SA	0.981	**0.998**	**0.982**	**0.982**	0.974	**0.984**	1.000
Ours	0.970	0.987	0.936	0.962	0.951	0.964	0.985

Table III
THE TOP-1 ACCURACY OF THE PROPOSED NETWORK AND COMPARE IT TO RECENT APPROACHES ON THE ISIC 2019 DATASET.

Method	Method in [24]			Ours		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
AK	91.4	48.4	96.5	91.9	54.6	96.1
BCC	94.9	72.1	94.0	94.1	73.2	93.6
BKL	90.4	39.4	98.5	91.4	42.6	98.0
DF	97.9	57.8	99.2	97.7	58.2	98.1
MEL	92.8	59.4	96.2	92.4	60.1	95.8
NV	96.0	71.0	97.5	96.5	71.6	97.1
VASC	95.6	64.4	99.1	95.1	64.6	98.5
SCC	93.8	43.9	98.6	92.2	48.3	98.3
UNK	77.5	0.3	99.9	78.2	38.5	97.8

conduct a thorough analysis, and compare the performance of our framework against state-of-the-art approaches.

1. Dataset

The experiments are performed on three separate datasets including the HAM dataset [25], the ISIC 2017 dataset, and the ISIC 2019 dataset.

The HAM10000 dataset [25] consists of 10,015 dermatoscopy images, each sized 450×600. It encompasses seven distinct diagnostic categories: AK, DF, BCC, MEL, VASC, BKL, and NV. Before their inclusion in the predictive model, all samples are reshaped to dimensions of $224 \times 224 \times 3$. Within the dataset, 9,187 images are allocated for the training set, and the testing set comprises 828 images.

The ISIC 2017 dataset comprises 2,600 images, each with dimensions of 767×1022. Within the training dataset, a total of 2,000 images representing three distinct diseases: benign tumors, seborrheic keratosis, and melanoma. Complementing this, the testing dataset encompasses 600 images for evaluation purposes.

The ISIC 2019 dataset represents the most expansive compilation of publicly available, quality-assured dermatoscopy images designed for research endeavors. This all-encompassing dataset encompasses a total of 25,331 images meticulously categorized into nine distinct classes. An illustrative example featuring three datasets is presented in Fig. 1, where the images from the HAM10000 dataset are showcased at the top, those from the ISIC 2017 dataset are displayed in the middle, and the example images from the ISIC 2019 dataset are depicted at the bottom.

2. Impremention Detail

Within the paper, we undertake both over-sampling and under-sampling techniques to attain an equitably distributed number of images across each class. Subsequently, the images undergo normalization by scaling each pixel through division by 255, thereby constraining pixel values within the range of 0 to 1. Our chosen batch size is set at 128. Employing the Stochastic Gradient Descent (SGD) optimizer, we initialize the learning rate to 0.01, which is subject to a 10-fold reduction if, over 10 consecutive epochs, the model fails to exhibit performance enhancements. Notably, the data augmentation strategies adopted are rooted in the RandAugment methodology [26]. All evaluations of test results are conducted using the dedicated test dataset. All experiments have been conducted in the PC with dual CPU E5-2673 v3 2.4Ghz, 64GB of RAM, and a single GPU A4000.

3. Performance and Comparison

First, we compare the proposed network with other recent methods such as [7, 21–23]. The results have shown in Table II, we found that the proposed framework archives competitive performance on all classes including AKIEC, BKL, DF, MEL, and VASC. Specifically, we achieve 97.0% in terms of accuracy on the class AKIEC outperforms [22] using VGG network, 98.7% in terms of accuracy on the class BCC outperforms [22] using VGG network, 96.2% in terms of accuracy on the class DF outperforms [21] using DenseNet model and so on. Although our network is not the best model, it is the smallest network i.e., using fewer parameters and computational costs compared to very deep networks such as VGG Net, DenseNet, ResNet, and SENet.

Subsequently, we assess the effectiveness of the proposed network by conducting a comparative analysis with established methodologies, including [7, 16, 27, 28], using the ISIC 2017 dataset. The results presented in Table IV demonstrate the superior performance of our proposed method across various metrics. In particular, the proposed network archives 92.5%, 87.8%, 90.2%, and 70.2% in terms of AUC, accuracy, sensitivity, and specificity, respectively. Notably, our network secures a top-2 position, trailing only behind Irv2+SA [17], in terms of sensitivity and accuracy.

Table IV
THE TOP-1 ACCURACY OF THE PROPOSED NETWORK AND COMPARED TO RECENT MODELS ON THE ISIC 2017 DATASET.

Model	AUC	Accuracy	Sensitivity	Specificity
ResNet-50 [7]	0.948	0.842	0.867	0.837
RAN50 [27]	0.942	0.862	0.878	0.859
SENet50 [28]	0.952	0.863	0.856	0.865
ARL-CNN50 [16]	**0.958**	0.868	0.878	**0.867**
Irv2+SA [17]	0.935	**0.898**	**0.945**	0.711
Ours	0.925	0.878	0.902	0.702

We extend the experiment on the ISIC 2019 dataset and compare our framework with the method in [24]. Table III shows that the proposed framework gives better results than [24] in most classes in terms of standard measures such as AUC, sensitivity, and specificity. Especially, our proposed framework achieves 91.9%, 91.4%, 96.5%, and 78.2% on AK, BKL, NV, and UNK, respectively outperforms the method in [24] by a margin of 0.5%–1%.

Finally, we conduct a comparison of the model size and computational cost, i.e., FLOPs, of our framework with other recent methods for SLC. The data in Table V indicates that our network achieves the smallest size and the lowest FLOPs when compared to various methods such as [17, 21, 23]. Specifically, our model utilizes only 0.2 million parameters and 0.3 GFLOPs for SLC, surpassing VGGNet by more than 600 times in terms of model size and 26 times in terms of computational cost. In comparison to the Irv2+SA method [17], our framework reduces the model size by 120 times and FLOPs by 20 times. These findings underscore that our framework significantly alleviates the memory requirements and inference time of the model, facilitating easy deployment on embedded and/or mobile devices without the necessity for GPU support.

Table V
THE DETAIL ABOUT THE NUMBERS OF PARAMETERS AND COMPUTATIONAL COST (GFLOPS) AMONG OUR PROPOSED FRAMEWORK AND RECENT APPROACHES. NOTE THAT 1 GFLOPS EQUALS 10^9 FLOPS.

Method	Architecture	Model size	FLOPs
Inception-v4 [23]	InceptionNet	7M	2G
DenseNet [21]	DenseNet	8M	3G
VGGNet [22]	VGGNet	133M	8G
ResNet-50 [7]	ResNet	45M	8G
SENet50 [28]	SENet	1M	0.8G
Irv2+SA [17]	Inception ResNet v2	24M	6G
Ours	**MobileNet**	**0.2M**	**0.3G**

IV. CONCLUSION

This paper presents the introduction of a lightweight network for skin lesion classification, leveraging MobileNet as its foundation. Rather than employing conventional convolution layers, our approach utilizes depthwise and pointwise convolutions to effectively reduce both the model size and computational cost of the network. Experimental results across diverse datasets demonstrate that our proposed framework attains competitive performance compared to recent methods, all the while utilizing only 0.2M parameters and 0.3G FLOPs. Moving forward, our research endeavors will focus on further analysis and the development of new efficient and lightweight networks to enhance performance, while concurrently minimizing memory size and inference time.

REFERENCES

[1] T. S. C. Foundation, "Skin cancer facts and statistics," https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/, access: 2023-06-30.
[2] M. UK, "2020 melanoma skin cancer report," https://www.melanomauk.org.uk/2020-melanoma-skin-cancer-report, access: 2023-06-30.
[3] R. K. Voss, T. N. Woods, K. D. Cromwell, K. C. Nelson, and J. N. Cormier, "Improving outcomes in patients with melanoma: strategies to ensure an early diagnosis," *Patient related outcome measures*, pp. 229–242, 2015.
[4] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The lancet oncology*, vol. 3, no. 3, pp. 159–165, 2002.

[5] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, S. Fröhling *et al.*, "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *European Journal of Cancer*, vol. 111, pp. 148–154, 2019.

[6] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, T. Holland-Letz *et al.*, "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *European Journal of Cancer*, vol. 113, pp. 47–54, 2019.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[8] C. Nhan Duong, K. Luu, K. Gia Quach, and T. D. Bui, "Longitudinal face modeling via temporal deep restricted boltzmann machines," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5772–5780.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[10] D.-Q. Vu, N. Le, and J.-C. Wang, "Teaching yourself: A self-knowledge distillation approach to action recognition," *IEEE Access*, vol. 9, pp. 105 711–105 723, 2021.

[11] Q. V. Duc, T. Phung, M. Nguyen, B. Y. Nguyen, and T. H. Nguyen, "Self-knowledge distillation: an efficient approach for falling detection," in *International Conference on Artificial Intelligence and Big Data in Digital Era*. Springer, 2021, pp. 369–380.

[12] H. M. Tan, D.-Q. Vu, C.-T. Lee, Y.-H. Li, and J.-C. Wang, "Selective mutual learning: An efficient approach for single channel speech separation," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3678–3682.

[13] H. M. Tan, D.-Q. Vu, and J.-C. Wang, "Selinet: A lightweight model for single channel speech separation," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[14] T. Phung, V. T. Nguyen, T. H. T. Ma, and Q. V. Duc, "A (2+ 1) d attention convolutional neural network for video prediction," in *International Conference on Artificial Intelligence and Big Data in Digital Era*. Springer, 2021, pp. 395–406.

[15] D.-Q. Vu and T. P. T. Thu, "Simultaneous context and motion learning in video prediction," *Signal, Image and Video Processing*, pp. 1–10, 2023.

[16] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2092–2103, 2019.

[17] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*. Springer, 2021, pp. 13–23.

[18] D. M. Nguyen, T. T. Nguyen, H. Vu, Q. Pham, M.-D. Nguyen, B. T. Nguyen, and D. Sonntag, "Tatl: Task agnostic transfer learning for skin attributes detection," *Medical Image Analysis*, vol. 78, p. 102359, 2022.

[19] A. T. Huynh, V.-D. Hoang, S. Vu, T. T. Le, and H. D. Nguyen, "Skin cancer classification using different backbones of convolutional neural networks," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2022, pp. 160–172.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[24] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," *MethodsX*, vol. 7, p. 100864, 2020.

[25] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[26] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 702–703.

[27] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

To Huu Nguyen received the Bachelor's Education of Information Technology at Thai Nguyen University of Education in 2003 and Master's degree in Computer Science at Thainguyen University in 2008. He received the PhD degree at the Academy of Science and Technology - Vietnam Academy of Sciences in 2021. He worked as a lecturer at the Faculty of Information Technology, School of Information and Communication Technology, Thai Nguyen University from 2004. Now, he is a researcher at the Institute of Information Technology, Academic Institute of Science and Technology, Vietnam.

Email: thnguyen@ictu.edu.vn

Ma T. Hong Thu received a master's degree in Computer Science in 2015 at the Thai Nguyen University of Information and Communications Technology. Research interests include: Machine learning and deep learning.
Email: thutq7@gmail.com

Do Thanh Mai received the Bachelor of Education in Information Technology at Thai Nguyen University of Education in 2003 and the Master's degree in Computer Science at Thai Nguyen University in 2008. She worked as a lecturer of Information Technology in the Basic Sciences Department at the School of Foreign Languages (SFL-TNU). Office address: School of Foreign Languages, Thai Nguyen University, Thai Nguyen, Vietnam.
Email: dothanhmai.sfl@tnu.edu.vn

Trang Phung T. Thu was born in Bac Ninh, Vietnam in 1991. She received a B.S. degree in education in information technology from the Thai Nguyen University of Education, Vietnam, in 2013 and an M.S. degree from the Thai Nguyen University of Information and Communication Technology (ICTU) in 2015.

Her research interests include machine learning, deep learning, computer vision, speech processing, and bioinformatics.
Email: phungthutrang.sfl@tnu.edu.vn

An Dang received her Ph.D. in Computer Science and Information Engineering from National Central University, Taiwan in 2022. She specializes in deep learning applications for image processing and audio/speech signal analysis.
Email: an.dangthithuy@phenikaa-uni.edu.vn

Duc-Quang Vu was born in Nam Dinh, Vietnam in 1991. He received a B.S. degree in education in information technology from the Thai Nguyen University of Education, Vietnam, in 2013 and an M.S. degree in Information Systems, from the University of Engineering and Technology (UET), Vietnam National University, Hanoi (VNU) in 2016. He received the Ph.D. degree in the Department of Computer Science & Information Engineering, National Central University, Taiwan in 2022 and a postdoc in 2023. His research interests include machine learning, deep learning, computer vision, speech processing, and bioinformatics.
Email: quangvd@tnue.edu.vn