

Optimizing DenseNet for Image Classification: A Synergy of Dual-Adaptive Attention and Multi-Scale Feature Extraction

Butaev Khusan Tashtemirovich*, Heejoung Hwang**

Abstract

In this paper, we introduce a new optimization method for DenseNet. In general, the most difficult part of medical image processing is the difficulty of improving performance for certain areas, such as images of skin, and we complemented this with a multi-scale feature extraction module that captures details at various resolutions. To ensure comprehensive learning, we used an adaptive loss function and a lite moderated channel block method that fine-tuned the model with detailed mathematical corrections to optimize sensitivity and specificity. Through experiments, we were able to confirm the excellence of the model as a result of extensive testing on complex data sets, and achieved an accuracy of 96.67%, proving that it is an efficient and reliable DenseNet model. This is a superior result not only to the general DenseNet model but also to other famous models such as MobileNet and VGG19, showing that it can be used in specific areas targeting skin images.

요약

본 논문에서는 DenseNet을 위한 새로운 최적화 방법을 소개한다. 일반적으로 의료 이미지 프로세싱에서 가장 어려운 부분은 특정 분야, 예를 들면 피부와 같은 이미지에 대해 성능을 향상시키기 힘들다는 것이고 우리는 다양한 해상도에서 세부 정보를 캡처하는 다중 규모 특징 추출 모듈을 통해 보완 했다. 포괄적인 학습을 보장하기 위해 적응형 손실 기능을 활용하고 감도와 특이성을 최적화 하기 위해 세밀한 수학적 보정으로 모델을 미세 조정하는 lite modified channel block 방식을 사용 하였다. 실험을 통해 복잡한 데이터 셋트에 대한 광범위한 테스트결과 모델의 우수성을 확인할 수 있었으며 96.67%의 정확도를 달성하여 효율적이면서도 신뢰할만한 DenseNet 모델임을 증명 하였다. 이는 일반 DenseNet 모델 뿐만 아니라 MobileNet, VGG19등 다른 유명 모델들에 비해서도 우수한 결과로 피부 이미지를 대상으로 하는 특정 영역에서의 활용이 가능함을 보여주었다.

Keywords

skin lesion classification, DenseNet, attention mechanism. modified classification model, class imbalance challenge

1. Introduction

Skin lesions, particularly melanoma or cancer, represent significant health risks. Medical imaging

enhances patient care by delivering intricate details on the body's internal structures and organs. Such insights are crucial for informing surgical approaches and formulating treatment strategies, contributing to

* Dept. of IT Convergence Engineering, Gachon Univ.
- ORCID: <https://orcid.org/0009-0006-2740-2423>
** Dept. of Computer Engineering, Gachon Univ.
- ORCID: <https://orcid.org/0000-0001-5462-0983>

• Received: Apr. 30, 2024, Revised: Jun. 20, 2024, Accepted: Jun. 23, 2024
• Corresponding Author: Heejoung Hwang
Dept. of Computer Engineering, Gachon Univ. Korea
Tel.: +82-31-750-4758, Email: hwanghj@gachon.ac.kr

improved patient results[1]. As a pivotal component of the healthcare infrastructure, medical imaging offers indispensable data for diagnosing, planning treatments, and tracking therapy progress[2]. This task holds significant importance in dermatology for facilitating early detection and diagnosis of skin cancer. A prevalent method for classifying skin lesions involves deep learning, utilizing deep neural networks like Convolutional Neural Networks(CNNs). These models are trained on extensive datasets of labeled skin lesion images, enabling them to extract and learn intricate features from the images to enhance classification precision[3].

Traditional image classification methods relied on features extracted through manually designed algorithms or statistical techniques. CNNs have seen swift advancements in various visual tasks including image classification. Beginning with the development of AlexNet[4], followed by VGG[5], and then GoogleNet[6], the performance of convolutional neural networks has improved with the increasing depth of these models. The development of GoogleNet structures[6], has been instrumental in consistently enhancing model accuracy while simultaneously reducing computational expenses. ResNet[7], revolutionized the training of deep convolutional neural networks through its introduction of residual connections. These connections allow gradients to flow through the network more effectively, enabling the training of networks that are much deeper than was previously possible without suffering from vanishing or exploding gradients. Following this, ResNeXt[8], further enhanced network performance by incorporating group convolution. MobileNet[9], designed for efficiency on mobile devices, trades off accuracy for speed, making it less precise than more complex models. Fine-tuning its parameters for optimal performance is also challenging. While InceptionResNet[10] combines the advantages of Inception and ResNet to handle complex tasks with high efficiency, it also inherits some challenges. The

EfficientNetB0 model[11], despite its efficiency and accuracy in various tasks, encounters limitations in specific scenarios. It is computationally demanding for deployment on devices with limited processing capabilities, such as smartphones or embedded systems. DenseNet[12], achieves dense connectivity and feature reuse along the channel dimension by concatenating the outputs of all preceding layers to serve as inputs for subsequent layers.

The main goal of this paper is to enhance skin lesion classification performance by advancing the capabilities of DenseNet. This involves innovating upon the DenseNet architecture to improve its effectiveness in image classification tasks, aiming to achieve higher accuracy and efficiency in processing and analyzing images. The paper introduces a novel model for advanced image classification by integrating modified dense blocks with spatial and channel attention mechanisms, specifically designed to enhance feature extraction, attention to relevant image regions, and overall classification performance. By addressing the limitations of existing models in handling complex visual tasks, this work contributes to the field through improved accuracy and efficiency in image classification tasks, offering significant advancements over traditional CNN-based approaches[13]. This approach demonstrates superior performance in various datasets, marking a step forward in the development of more sophisticated and effective deep-learning models for image analysis.

II. Related Works

The integration of attention mechanisms into CNNs, specifically channel and spatial attention, represents a pivotal advancement in enhancing image classification and feature extraction. These mechanisms enable models to selectively concentrate on pertinent semantic details, thereby refining their interpretability and accuracy. DenseNet, a foundational architecture in this domain, introduces dense connections that facilitate

efficient feature reuse and minimize information loss, setting a benchmark for subsequent innovations.

2.1 Skin lesion classification

The classification of skin cancer poses significant challenges due to the presence of artifacts, variations in image resolution, and subtle differences in features between various cancer types. Clinical procedures for examining skin lesions are often perceived by patients as complex and uncomfortable and can fall short of effectively differentiating between lesion categories. Advanced techniques in Machine Learning and Computer Vision present a promising avenue to address the hurdles associated with skin lesion classification. The foundational step in Computer-Aided Diagnosis(CAD) systems involves generating discriminative features to categorize lesions as normal or pathological. The classification of skin cancer presents considerable challenges due to the presence of artifacts, varying image resolutions, and subtle distinguishing features among different cancer types. Such complexities render clinical procedures for skin lesion assessment both intricate and oftentimes ineffectual in accurately differentiating between lesion categories.

In response to these diagnostic challenges, we have crafted a novel model based on the DenseNet architecture[14]. Over the past several years, significant strides have been made to support medical professionals in the accurate recognition of skin lesions. The development of CAD systems for skin abnormalities has become a rapidly evolving area of research. When it comes to the categorization of images, there exist two predominant techniques: conventional machine learning methods and the more recent advancements in deep learning.

2.1.1 Traditional machine learning methods

In recent years, considerable initiatives have been

undertaken to support medical professionals in the accurate recognition of skin lesions. Image classification within these systems is typically approached via two distinct methodologies: conventional machine learning and the more advanced deep learning techniques. The study by [14] integrated gradients and Local Binary Patterns(LBP) to delineate the border characteristics of lesion segmentation. For classification, three methods were utilized: SVM - SMO(Sequential Minimal Optimization), SVM - ISDA (Iterative Single Data Algorithm), and Feedforward Neural Network(FNN). The research also incorporated two additional datasets, Dermoft and MED-NODE, to evaluate performance. The results on the MED-NODE dataset revealed specificity, accuracy, and sensitivity rates of 88%, 79%, and 65%, respectively. Despite the advancements in machine learning techniques, their effectiveness in the field of skin lesion classification remains somewhat limited, often yielding less than optimal accuracy. These methods come with several limitations, including challenges in achieving high precision in diagnoses.

2.1.2 Deep learning methods

Due to significant limitations encountered with machine learning approaches, particularly in achieving satisfactory accuracy, researchers have increasingly turned to deep learning techniques for the classification of skin lesions[15].

Numerous studies have been conducted to classify skin lesions utilizing various datasets. One such study employed ResNet18 with different optimizers and extracted skin image features from the ISIC2017 dataset. The effectiveness of this method was evaluated using the ISIC2017 dataset, with ResNet50 yielding an accuracy of 93.7%, specificity of 98.3%, and sensitivity of 83.3%. Further research has explored the integration of pre-trained deep networks, such as ResNet, AlexNet, GoogleNet, and VGG, with transfer learning to enhance classification performance. A

system utilizing a Deep Convolutional Neural Network (DCNN) combined with the Fisher vector for recognizing and classifying dermoscopy images was proposed by [16]. Additionally, [17] implemented a method that extracts and integrates features of skin lesions based on CNN and the ABCD rule.

The classification of these features was carried out using the Relevance Vector Machine(RVM), linear regression(LR), and Support Vector Machine (SVM), achieving a classification rate of 89.71%. [16] enhanced a melanoma detection and classification method that merges feature extraction through Deep Convolutional Neural Network(DCNN) with Linear Discriminant Analysis(LDA), achieving a classification accuracy of 86%. [19] introduced a melanoma classification system where lesions were segmented utilizing the Convergence of Intermediate Decaying OmniGradients(SCIDOG) method alongside Synthesis. The Regions of Interest(ROI) were classified using the Predict-Evaluate-Correct K-fold(PECK) technique, which integrates DCNN with Random Forest(RF) and Support Vector Machine(SVM), resulting in a categorization success rate of 91%.

The aforementioned methods, while advancing melanoma detection and classification through innovative combinations of Deep Convolutional Neural Network(DCNN) feature extraction, machine learning algorithms, and segmentation techniques, exhibit certain limitations. The integration of DCNN with Linear Discriminant Analysis(LDA) and the novel segmentation using SCIDOG followed by classification with the PECK method, although innovative, suggest that further refinement and exploration of advanced techniques are essential to enhance performance. These limitations underline the ongoing need for research to address the intricacies of skin lesion classification, aiming for higher accuracy, specificity, and sensitivity in detecting and classifying melanoma and other skin lesions.

III. Proposed Method

Our proposed method integrates dense blocks with attention mechanisms, leveraging their strengths in image classification to enhance feature extraction and model performance. Dense blocks facilitate efficient information flow and feature reuse, while attention mechanisms allow the model to focus on the most relevant parts of the image, improving accuracy and interpretability.

3.1 Modified dense blocks

Fig. 1 shows that the proposed model consists of dense blocks from the image classification model DenseNet. The model includes three dense blocks and attention mechanisms, respectively. Dense blocks are distinguished by their dense connectivity patterns among layers within the block. In contrast to conventional architectures with sequential layer connections, each layer in a dense block receives direct input from all preceding layers.

This dense connectivity promotes feature reuse, facilitates gradient flow during training, and enhances the overall effectiveness of the network. This choice is motivated by the fact that the bottleneck section of the DenseNet utilizes an input feature size of 64, while the initial input image possesses a shape of $3 \times 32 \times 32$ with 3 input features.

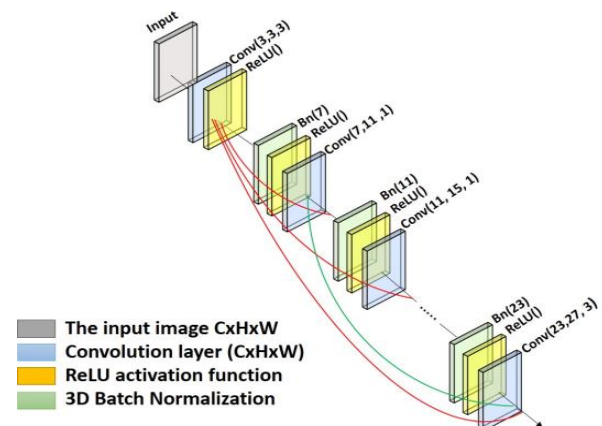


Fig. 1. Modified 6-layer dense block with growth rate of $k=4$ of the proposed model

Fig. 1 illustrates the modified dense block, emphasizing enhanced connectivity and efficient feature reuse within the network. The diagram uses different colors to represent the components: gray for the input image (CxHxW), blue for convolution layers, yellow for ReLU activation functions, and green for 3D Batch Normalization. This visual representation clarifies the data flow and interactions between layers, aiding in the understanding of the model's structure and performance improvements.

The utilization of a convolution layer allows an increase in the number of input features by 64, addressing the potential inadequacy of input features if only the original 3 channels were used. The input image, denoted as x_0 , traverses through a sequence of l dense and attention blocks, generating outputs at each block for the l -th block in the sequence.

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

where $[x_0, x_1, \dots, x_{l-1}]$ denotes the amalgamation of feature-maps generated in layers $[0, \dots, l-1]$. The term DenseNet is attributed to this network architecture due to its dense connectivity. To facilitate implementation, we streamline the multiple inputs of H_l in equation (1) by combining them into a unified tensor. If each function H_l generates k feature-maps, then it can be deduced that the l -th layer encompasses $k_0 + k \times (l-1)$ input feature-maps, where k_0 represents the number of channels in the input layer.

Here, the hyperparameter k is denoted as the growth rate of the network, in our proposed model we use $k = 4$. The feature maps essentially represent the global state of the network, and each layer augments this state by adding k feature maps. The growth rate serves as a regulatory factor, determining the extent to which each layer contributes new information to the global state. Though each layer typically yields only k output feature maps, it frequently accommodates a more extensive array of inputs. It has been observed in prior studies that the incorporation of a 1×1

convolution as a bottleneck layer preceding each 3×3 convolution proves effective in diminishing the number of input feature maps, thereby enhancing computational efficiency.

The bottleneck layer is represented as BN-ReLU-Conv (1×1) BN-ReLU-Conv (3×3) in version H_l . In our experimental setups, we assign each 1×1 convolution the task of producing growth rate k feature maps.

3.2 Effectiveness of the light-modified channel attention module

In the realm of CNNs, attention modules play a crucial role in enhancing learning by directing the model to concentrate on pertinent information while disregarding irrelevant background details. In the context of object detection, the salient data comprises the cropped representation of objects or target classes that necessitate classification and localization within an image. The Convolutional Block Attention Module (CBAM) segregates into dual submodules: channel attention and spatial attention. Each submodule incorporates convolutional strata complemented by maximal and average pooling procedures.

The input feature F , a product of the preceding modified dense block, undergoes a fusion of Max and Avg poolings, yielding two distinct spatial context descriptors, F_{avg}^c and F_{max}^c . Subsequently, these descriptors are propagated through a Multi-layer Perceptron (MLP) equipped with a singular hidden layer, culminating in a feature representation $M_c \in R^{c \times 1 \times 1}$.

To curtail parameter proliferation, the MLP condenses its size to $\left(\frac{c}{r}\right)^{1 \times 1}$ per output feature, with r denoting the reduction ratio. Following the application of the shared network to each descriptor, the resultant feature vectors are amalgamated via an element-wise additive process.

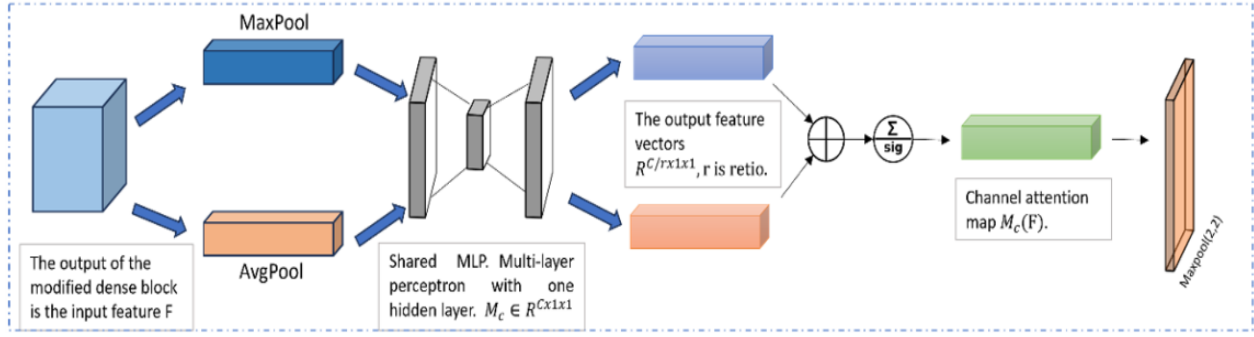


Fig. 2. Light-modified channel attention block

In this study, we preserve the integrity of the Channel Attention Block(ChaB), noting its previously established efficacy, and introduce only Maxpooling to diminish the vector map dimensions and enhance informational retrieval.

Fig. 2 the CBAM is fed an intermediate feature map $F_{MDB} \in R^{CxHxW}$, emanating from a Modified Dense Block (MDB), methodically deduces two forms of attention maps. The first is a 1D channel attention $M_c \in R^{CxHxW}$ and the second is a 2D spatial attention map, $M_s \in R^{CxHxW}$, as delineated in Equation (2).

$$\begin{aligned} F' &= M_c(F) \otimes F_a \\ F' &= M_c(F') \otimes F \end{aligned} \quad (2)$$

Given the emphasis on the classification task, the channel attention module, which is adept at discerning 'what' aspects are present within the input feature, is utilized. Consequently, only $M_c(F)$ is extracted and integrated with the output of the i -th MDB as per Equation (3). This selective application ensures that the focus remains on channel-wise feature relevance, pertinent to identifying class-discriminative information.

$$F'(F_{MDB_i}) = M_c(F_{MDB_i}) \otimes F_{MDB_i} \quad (3)$$

The intire process of Light-modified ChaB can be encapsulated by Equation (4) where σ represents the sigmoid activation function. The weights of the i -th

layer of the MLP, denoted as $W_i \in R^{\left(\frac{c}{r}\right) \times c}$, and the weights of the preceding layer, $W_i \in R^{c \times \left(\frac{c}{r}\right)}$, are specifically designed to be shared across both inputs, facilitating a consistent parameterization for the dual pathways involving MaxPooling and AveragePooling. Following this parameter sharing and the application of the ReLU activation function, a MaxPooling operation is employed as delineated in Equation (5). This incorporation of MaxPooling constitutes a subtle yet impactful modification aimed at optimizing the feature selection process within the channel attention framework, enhancing the module's efficiency in focusing on pivotal aspects pertinent to the classification task at hand.

$$\begin{aligned} M_c(F_{MDB_i}) &= \sigma(MLP(AvgPool(F_{MDB_i}))) + \\ &MLP(AvgPool(F_{MDB_i})) = \\ &\sigma(W_i(W_{i-1}(F_{avg}^c)) + W_i(W_{i-1}(F_{max}^c))) \end{aligned} \quad (4)$$

$$M_c(F_{MDB_i})' = MaxPool(M_c(F_{MDB_i})) \quad (5)$$

Combine the outcomes of average pooling and max pooling and the extra maxpoling, we transmit the concatenated result through a compact convolutional block featuring a 3x3 kernel size. As a consequence, the output of Light-modified ChaB proceeds to theto the next i -th MDB. The dense block and attention block feature extraction cycle is repeated about three

times, then the output goes into flatten layer and the final output comes from activation function softmax, that takes as input a vector of numerical scores or logits and normalizes them into a probability distribution, as shown:

$$Z_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (6)$$

where z_i input vector of each element, by raising e to the power of the input vector and dividing the result by the sum of the exponentiated values across all elements in the vector.

IV. Experimental Results

This section offers a detailed exposition of the methodologies employed for data preprocessing and the configuration of the experimental framework.

4.1 Dataset

The International Skin Imaging Collaboration (ISIC) 2018 dataset[20] represents a pivotal resource in the domain of dermatological image analysis, specifically focusing on the automated detection and classification of skin lesions. Assembled for the ISIC Skin Lesion Analysis Towards Melanoma Detection Challenge, this comprehensive dataset encompasses a total of 37,649 dermoscopic images.

4.2 Results of the proposed model

Table 1 presents a comparative analysis of various classification networks when evaluate on the ISIC dataset. The networks compare include DenseNet121, VGG19, MobileNet, EfficientNetB0, ResNet50, and a the proposed model. For DenseNet121, the training accuracy is recorded at 96.56%, with a corresponding validation accuracy of 95.70%. It shows a training loss of 0.073 and a validation loss of 0.124 over 60 epochs, leading to a test accuracy of 95.94%. The model has 8 million parameters.

Furthermore, the performance of the VGG19 shows a training accuracy of about 91.23% and a validation accuracy of 88.56%, with higher loss values of 0.279 for training and 0.342 for validation. Despite 60 epochs of training, the test accuracy is lower than DenseNet121, at 90.27%, and the model has a significantly larger number of parameters, amounting to 21 million.

MobileNet achieves a training accuracy of 96.87% and a validation accuracy of 94.34%, with training and validation losses of 0.063 and 0.145, respectively. After 60 epochs, the test accuracy is 93.43%, and the model is relatively efficient with 4 million parameters. EfficientNetB0 reports training and validation accuracies of 96.78% and 94.67%, respectively, with the lowest training loss of 0.025 and a validation loss of 0.213.

EfficientNetB0, the model with the least parameters at 6 million, leads with the highest training and validation accuracies of 96.78% and 95.14%, respectively.

Table 1. Classification with different models

Classification models	Train accuracy	Validation accuracy	Train loss	Validation loss	Epochs	Test accuracy	Parameters
DenseNet121	96.56	95.70	0.073	0.124	60	95.94	8M
VGG19	91.23	88.56	0.279	0.342	60	90.27	21M
MobileNet	96.87	94.34	0.063	0.145	60	93.43	4M
EfficientNetB0	96.78	94.67	0.025	0.213	60	95.14	6M
ResNet50	89.84	86.26	0.321	0.632	60	87.47	25.6M
The proposed model	97.12	95.84	0.043	0.082	60	96.67	27M

ResNet50, with 25.6 million parameters, has training and validation accuracies of 89.84% and 75.15%. ResNet50 shows a training accuracy of 87.44% and a validation accuracy of 86.26%, with higher losses (0.321 for training and 0.632 for validation).

V. Conclusion

This paper presents a detailed exploration of a novel approach for image classification, leveraging the synergy of dense blocks and attention mechanisms. By incorporating these two powerful components, we have developed a model that significantly enhanced feature extraction capabilities and improves classification accuracy. The methodology addresses critical challenges in image classification, providing a sophisticated solution that combines efficiency, precision, and adaptability. Our results affirm the effectiveness of our proposed framework, paving the way for future advancements in deep learning applications. This work not only demonstrates the potential of integrating dense connectivity and attention mechanisms but also sets a new benchmark for image classification tasks.

VI. Discussion

The proposed model has demonstrated exceptional performance in skin lesion classification tasks. In our experiments, the proposed model achieved a training accuracy of 97.12% and a validation accuracy of 95.84%, which are superior to those of DenseNet121, VGG19, MobileNet, EfficientNetB0, and ResNet50. DenseNet121, while a strong performer with a training accuracy of 96.56% and validation accuracy of 95.70%, falls short of the proposed model's results. VGG19, despite its deep architecture, shows a lower validation accuracy of 88.56% and a higher parameter count of 21 million, making it less efficient compared to the proposed model.

MobileNet, known for its efficiency on mobile devices, achieved a validation accuracy of 94.34%, but its trade-off between speed and accuracy results in a lower performance compared to our proposed model. EfficientNetB0, with its impressive architecture, shows a validation accuracy of 94.67%, yet it still does not match the proposed model's validation accuracy of 95.84%. Additionally, EfficientNetB0, despite having fewer parameters (6 million), cannot outperform the proposed model in terms of accuracy. ResNet50, with a substantial parameter count of 25.6 million, demonstrates lower validation accuracy (86.26%) and higher validation loss (0.632) compared to the proposed model. The proposed model not only surpasses ResNet50 in accuracy but also maintains a lower validation loss of 0.082.

The proposed model use of dual-adaptive attention mechanisms and multi-scale feature extraction significantly contributes to its superior performance. These enhancements allow the model to focus more effectively on relevant features and details within the images, leading to higher classification accuracy. Moreover, the integration of modified dense blocks ensures efficient feature reuse and improved gradient flow, further boosting the model effectiveness.

While the proposed model has a higher parameter count (27 million) compared to DenseNet121 (8 million), its performance gains justify this increase. Future work will focus on optimizing the model to reduce parameters without compromising performance, ensuring that it remains both powerful and efficient.

References

- [1] X. Deng, "LSNet: a deep learning based method for skin lesion classification using limited samples and transfer learning", *Multimedia Tools and Applications*, Vol. 83, pp. 61469-61489, Jan. 2024. <https://doi.org/10.1007/s11042-023-17975-2>.
- [2] D. Chanda, M. S. H. Onim, H. Nyeem, T. B.

- Ovi, and S. S. Naba, "DCENSnet: A new deep convolutional ensemble network for skin cancer classification", *Biomedical Signal Processing and Control*, Vol. 89, pp. 105757, Mar. 2024. <https://doi.org/10.1016/j.bspc.2023.105757>.
- [3] S. Savaş, "Enhancing Disease Classification with Deep Learning: a Two-Stage Optimization Approach for Monkeypox and Similar Skin Lesion Diseases", *Journal of Imaging Informatics in Medicine*, Vol. 37, pp. 778-800, Jan. 2024. <https://doi.org/10.1007/s10278-023-00941-7>.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6, pp. 84-90, May 2017. <https://doi.org/10.1145/3065386>.
- [5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, Sep. 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
- [6] C. Szegedy, et al., "Going deeper with convolutions", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1-9. Jun. 2015. <https://doi.org/10.1109/cvpr.2015.7298594>.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, Jun. 2016. <https://doi.org/10.1109/cvpr.2016.90>.
- [8] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 1492-1500, Jul. 2017. <https://doi.org/10.1109/cvpr.2017.634>.
- [9] A. G. Howard, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861*, Apr. 2017. <https://doi.org/10.48550/arXiv.1704.04861>.
- [10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning", In *Proc. of the AAAI conference on artificial intelligence*, San Francisco, California USA, Vol. 31, No. 1, Feb. 2017. <https://doi.org/10.1609/aaai.v31i1.11231>.
- [11] M Tan and Q Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", In *International conference on machine learning*, Long Beach, California, USA, Vol. 97, pp. 6105-6114, Jun. 2019.
- [12] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 4700-4708, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.243>.
- [13] S. Muksimova, S. Umirzakova, S. Mardieva, and Y.-I. Cho, "Enhancing Medical Image Denoising with Innovative Teacher-Student Model-Based Approaches for Precision Diagnostics", *Sensors* 2023, Vol. 23, No. 23, pp. 9502, Nov. 2023. <https://doi.org/10.3390/s23239502>.
- [14] K. Balaji, V. Nirosha, S. Yallamandaiah, S. Karthik, V. S. Prasad, and G. Prathyusha, "DesU-NetAM: optimized DenseU-net with attention mechanism for hyperspectral image classification", *International Journal of Information Technology*, Vol. 15, No. 7, pp. 3761-3777, Aug. 2023. <https://doi.org/10.1007/s41870-023-01386-5>.
- [15] S. Umirzakova, S. Ahmad, S. Mardieva, S. Muksimova, and T. K. Whangbo, "Deep learning-driven diagnosis: A multi-task approach for segmenting stroke and Bell's palsy", *Pattern Recognition*, Vol. 144, Dec. 2023. <https://doi.org/10.1016/j.patcog.2023.109866>.
- [16] P. M. M. Pereira, et al., "Skin lesion classification enhancement using border-line

features-The melanoma vs nevus problem", Biomedical Signal Processing and Control, Vol. 57, pp. 101765, Mar. 2020. <https://doi.org/10.1016/j.bspc.2019.101765>.

- [17] H. Hu, Z. Chen, and Y. Xia, "Encoding Deep Residual Features into Fisher Vector for Skin Lesion Classification", 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, pp. 1843-1846, Jul. 2022. <https://doi.org/10.1109/embc48229.2022.9871597>.
- [18] F. P. Loss, et al, "Skin cancer diagnosis using NIR spectroscopy data of skin lesions in vivo using machine learning algorithms", arXiv preprint arXiv:2401.01200, Jan. 2024. <https://doi.org/10.48550/arXiv.2401.01200>.
- [19] A. E. Adeniyi, et al., "Comparative Study for Predicting Melanoma Skin Cancer Using Linear Discriminant Analysis (LDA) and Classification Algorithms", Artificial Intelligence, Data Science and Applications, Vol. 837, pp. 326-338, Mar. 2024. https://doi.org/10.1007/978-3-031-48465-0_42.
- [20] 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, Aug. 2018. <https://doi.org/10.1038/sdata.2018.161>.

Heejoung Hwang



2008. 2 : Ph.D. in Computer Engineering, Incheon University
2000. 9 ~ Present : Professor, Dept. of Computer Engineering, Gachon University
Research Interests : Software Engineering, Software Architecture, Smart Factory, Healthcare IT

Authors

Butaev Khusan Tashtemirovich



2022. 2 : Bachelor's Degree in Computer Engineering, Gachon University
2022. 9 ~ Present : Master's Degree candidate, IT Convergence Engineering, Gachon University

Research Interests: AI, Deep Learning, Image Classification, Mobile App Framework