Development and Validation of Adaptable Skin Cancer Classification System Using Dynamically Expandable Representation

Bong Kyung Jang, Yu Rang Park
Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea

Objectives: Skin cancer is a prevalent type of malignancy, necessitating efficient diagnostic tools. This study aimed to develop an automated skin lesion classification model using the dynamically expandable representation (DER) incremental learning algorithm. This algorithm adapts to new data and expands its classification capabilities, with the goal of creating a scalable and efficient system for diagnosing skin cancer. Methods: The DER model with incremental learning was applied to the HAM10000 and ISIC 2019 datasets. Validation involved two steps: initially, training and evaluating the HAM10000 dataset against a fixed ResNet-50; subsequently, performing external validation of the trained model using the ISIC 2019 dataset. The model’s performance was assessed using precision, recall, the F1-score, and area under the precision-recall curve. Results: The developed skin lesion classification model demonstrated high accuracy and reliability across various types of skin lesions, achieving a weighted-average precision, recall, and F1-score of 0.918, 0.808, and 0.847, respectively. The model’s discrimination performance was reflected in an average area under the curve (AUC) value of 0.943. Further external validation with the ISIC 2019 dataset confirmed the model’s effectiveness, as shown by an AUC of 0.911. Conclusions: This study presents an optimized skin lesion classification model based on the DER algorithm, which shows high performance in disease classification with the potential to expand its classification range. The model demonstrated robust results in external validation, indicating its adaptability to new disease classes.

Keywords: Dermoscopy, Skin Neoplasms, Deep Learning, Decision Support Systems, Clinical, Diagnostic Imaging

I. Introduction

Skin cancer is one of the most common types of human malignancies. Its diagnosis typically involves visual examination, dermoscopy, and histopathological analysis. However, due to the high volume of daily examinations and the complexity of disease detection, delays in diagnosis can occur. Existing classification systems, developed in previous studies, consist of deep learning models trained under fixed conditions and are limited to classifying a small number of diseases. As a result, there is a growing demand for efficient and rapid diagnostic tools. Machine learning and deep learning models, particularly those analyzing medical images, have the potential to effectively meet this need [1-3]. To train
Adaptable Classification of Skin Cancer

deep learning models for accurate disease diagnosis and analysis based on medical images, a substantial amount of data is required. Additionally, as new diseases are discovered or existing diseases exhibit variations, hospitals encounter an increasingly diverse array of disease types. Consequently, classification models must be continually updated and expand their classification scope to remain relevant in clinical applications.

In the field of medical diagnostics, the development and implementation of computer-aided diagnosis (CAD) tools present significant challenges. One critical aspect is the need for incremental learning, especially when dealing with diverse pathologies and imaging machines. Creating a comprehensive dataset that includes various diseases and imaging modalities can be a time-consuming process, often taking several years to collect and annotate [4].

In this study, we propose a classification model that employs dynamically expandable representation (DER), a state-of-the-art incremental learning algorithm capable of scaling in the image classification domain [5]. The DER model addresses the challenges associated with the fine-grained variability in the appearance of skin cancers by continually learning and adapting to new data. It enables the efficient integration of new disease types into the classification framework, ensuring that the model remains up-to-date and capable of accurately classifying a wide range of skin cancers.

The objective of this study was to develop an automated skin lesion classification model using DER and to evaluate its performance. By leveraging the strengths of deep learning and incremental learning techniques, we aim to improve the efficiency and accuracy of skin cancer diagnosis. The model was trained on a large dataset of annotated skin lesion images, including diverse disease types, and its performance was evaluated using various metrics.

II. Methods

1. Data Construction

Two datasets were utilized in this study. The first, the HAM10000 (Human Against Machine with 10000 training images) dataset, included images categorized into distinct classes such as actinic keratoses, basal cell carcinoma, dermatofibroma, melanocytic nevi, melanoma, and vascular lesions [6]. The second dataset, International Skin Imaging Collaboration (ISIC) 2019, contained images classified into the same six categories as the first, with the addition of seborrheic keratosis and squamous cell carcinoma [7]. The datasets were divided into three subsets: training, validation, and test data, with the distribution set at 80%, 10%, and 10%, respectively. These datasets were used to train a model featuring an incremental learning structure. The effectiveness of the proposed method was subsequently evaluated.

2. Experimental Setting

All experimental settings were implemented using PyTorch 1.12.1. The models were trained on a machine equipped with four NVIDIA QUADRO P5000 GPUs, CUDA 11.2, 64 GB of memory, and an Intel Xeon Platinum 8253 CPU operating at 2.2 GHz.

3. Data Augmentation

To enhance the generalization capabilities of the convolutional neural network model, we employed extensive data augmentation techniques during the training phase. Initially, each skin image was resized to a dimension of 224 × 224 pixels to serve as the input for the model. We applied the following data augmentation steps:

- Horizontal flip: Each image was independently flipped horizontally with a probability of 0.5. This augmentation technique introduces variations in orientation, enhancing the model’s ability to process images with different spatial orientations.
- Random contrast and brightness change: Independent random adjustments were made to the contrast and brightness of each image. These adjustments were confined to a range of 20% in both the positive and negative directions and were applied with a probability of 0.5. This augmentation technique captures variations in illumination conditions within the dataset.
- Distortion: Distortion was applied to the images with a probability of 0.2. Various types of distortions, including optical, grid, and elastic distortion, were utilized. These techniques introduce geometric transformations to the images, simulating variations in image acquisition conditions and enhancing the model’s robustness to such variations.

These data augmentation techniques collectively contribute to expanding the diversity of the training dataset, enabling the model to learn more robust and generalized representations. By incorporating these variations into the training process, the model becomes better equipped to handle various image conditions and variations commonly encountered in real-world scenarios [8,9].

4. Incremental Learning

The DER algorithm for skin disease classification employs a multi-step approach, where each step involves training
and validation using data specific to that step. The features learned in previous steps are combined with those from the current step to create a super feature representation. This incremental learning process enables the model to expand its capability to classify a broader range of classes. In each step of the incremental learning process, an additional training loss, referred to as the auxiliary loss, is calculated. The auxiliary loss helps classify the classes from both previous and current steps. By incorporating the auxiliary loss, the model learns different features that facilitate the expansion of the classifiable classes [5] (Figure 1). This incremental learning strategy is crucial for adapting the model to handle an increasing number of disease classes over successive steps. The model progressively accumulates knowledge from previous steps while incorporating new information from the current step. As a result, the model becomes more versatile in its ability to accurately classify skin lesions across a broader spectrum of classes. By employing incremental learning, the DER model not only maintains its performance on previously learned classes but also extends its capabilities to new classes. This approach ensures that the model can continuously adapt to the evolving landscape of skin lesion classification, incorporating new knowledge and improving its diagnostic accuracy.

5. Study Design
In this study, we developed and validated a skin cancer classification model using the HAM10000 dataset along with an external validation dataset from ISIC 2019. The HAM10000 dataset comprises images of various skin cancer types, including actinic keratoses, basal cell carcinoma, dermatofibroma, melanocytic nevi, melanoma, and vascular lesions. The model underwent training in a stepwise manner, with each phase focusing on two specific classes and assessing its performance thereafter. The training process was divided into three steps, allowing the model to progressively learn and classify different types of skin lesions. In the first step, the model classified actinic keratoses and basal cell carcinoma. In the second step, it focused on dermatofibroma and melanocytic nevi. The third step involved training on melanoma and vascular lesions. To strengthen the model’s classification capabilities, we introduced an additional incremental step that involved training on external data. This new step incorporated seborrheic keratosis and squamous cell carcinoma from the ISIC 2019 dataset, which were not included in the initial three-step training process. For validation, we utilized the HAM10000 dataset to assess the model’s performance in comparison to a ResNet-50-based neural network model from a previous study [10]. Subsequently, we applied incremental learning to integrate new classes from the ISIC 2019 dataset, followed by external validation using the same dataset.

This study was designed to develop a robust and versatile model for classifying skin lesions, capable of accurately identifying a wide range of lesion types across various datasets.

6. Model Evaluation
In this study, we used specific indicators to evaluate deep learning algorithms in medical image classification. We compared the performance of the DER-based model with a deep convolutional neural network, known as ResNet-50, using the same dataset. The classification performance for skin cancer was assessed using precision, recall, F1-score, and the area under the precision-recall curve [11]. The evaluation process involved calculating these metrics for both the training and validation datasets.

In addition to the performance evaluation metrics, this study also utilized gradient-weighted class activation map-
Adaptable Classification of Skin Cancer

ping (Grad-CAM) to offer visual explanations for the skin lesions classification model [12]. Grad-CAM produces heat maps that emphasize the areas of the input images which are most influential in the model’s decision-making process. These heatmaps clearly illustrate the regions of interest and enhance the interpretability of the model’s classification decisions.

III. Results

In this study, we compared the performance of DER to that of ResNet-50 across various metrics using the HAM10000 dataset, which includes 1,015 images of skin cancer. The comparative analysis revealed that DER, an incremental learning algorithm, consistently outperformed the ResNet-50 model across all evaluated criteria (Table 1). The DER-based model achieved a classification accuracy of 80.88%, significantly exceeding the 73.55% accuracy of ResNet-50. Additionally, other performance metrics also demonstrated notable differences between the two models, particularly precision (weighted average precision: DER of 0.919, ResNet-50 of 0.799). Receiver operating characteristic (ROC) curve analysis further highlighted the model’s discrimination capability across different classification thresholds, with DER achieving an average area under the curve (AUC) value of 0.943 (Figure 2). The best-performing category was vascular lesions, with an AUC of 0.987, while the least effective was actinic keratoses, with an AUC of 0.888.

To further validate the model’s performance, external validation was conducted using the ISIC 2019 dataset. ROC curve analysis demonstrated the model’s ability to discriminate across different classification thresholds, achieving an AUC value of 0.911 (Figure 3). Consistent with the HAM10000 results, vascular lesions were the best-perform-
ing skin condition in the ISIC 2019 dataset, with an AUC of 0.992, while the model performed worst for actinic keratosis, with an AUC of 0.815. The Grad-CAM results, which provided a visual representation of the skin lesion classification model, showed that the model clearly detected skin lesions (Figure 4), regardless of other image characteristics such as variations in original skin color or the presence of hair.

IV. Discussion

The primary contribution of this study is the development of an optimized skin cancer classification model based on the DER algorithm [5]. This model facilitates the progressive expansion of the classification range for skin lesions. It was trained and validated using the HAM10000 dataset [6], which includes six different skin diseases. We utilized the same ResNet-50 structured neural network to compare the classification accuracy with that of a convolutional neural network model developed by training on fixed data at once in a previous study [10]. The DER-based model was trained incrementally and achieved a classification accuracy of 80.88%, surpassing the 73.55% performance of the ResNet-50. In a previous study [13], skin cancer classification models using various neural networks were compared. The results showed the following accuracies: Xception at 78.45%, DenseNet-201 at 78.59%, InceptionResNet-V2 at 79.39%, GoogLeNet at 79.45%, and AlexNet at 79.65%. These figures represent lower performance compared to the 80.88% achieved by the DER model proposed in this study.

The reason for DER’s superior performance over traditional models lies in its ability to retain and utilize information from previous samples to predict subsequent ones incrementally [14,15]. In contrast, traditional models do not recycle information across datasets; instead, they depend solely on the data within each dataset. This approach makes them particularly vulnerable to issues like small sample sizes.
Adaptable Classification of Skin Cancer

and non-independent identically distributed labels [16,17]. Through incremental learning, the model showed effectiveness in disease classification, as evidenced by an average AUC of 0.943 and a weighted average F1-score of 0.847. To further validate the model’s improved disease classification capabilities, external validation was performed using the ISIC 2019 dataset. During this phase, two new disease classes, seborrheic keratosis and squamous cell carcinoma, were incrementally trained and validated. The results from this external validation confirmed that the model successfully adapted to these new disease classes, achieving a competitive average AUC of 0.911. These findings underscore the model’s ability to expand its classification range and adapt to new disease types effectively. The incremental learning approach used in this study enabled the model to progressively accumulate knowledge from previous steps while integrating new information, thereby improving its ability to classify a diverse range of skin lesions. Additionally, the Grad-CAM images not only deepened our understanding of the model’s reasoning processes but also provided valuable insights for clinicians, enhancing their ability to assess the model’s reliability and verify its decision-making process [12].

We hypothesize that DER holds promise for fields that require scalability, particularly in medical screening and diagnostics. Dermatological conditions, including skin cancer, are likely to expand with the identification of new subtypes and broader diagnostic criteria. Given that the skin is one of the body’s largest organs, it provides a complex and diverse environment that is prone to disease manifestation [18]. The emergence of new environmental stressors, genetic variations, environmental pollutants, or changes in lifestyle habits may lead to the development of previously unknown skin diseases [19,20]. Moreover, advancements and refinements in medical technology offer opportunities to discover and accurately diagnose previously unrecognized dermatological conditions.

However, the developed model, while promising, has certain limitations and challenges that must be acknowledged. In this study, the use of imbalanced datasets for training led to variations in classification performance across different types of skin cancer. Further research is essential to determine the model’s generalizability to other datasets and its applicability in real-world clinical settings. Additionally, as new types of diseases emerge and existing ones evolve, continuous updates and adjustments to the model will be necessary.

In summary, this study implemented DER, an incremental learning method, to address the limitations associated with the resource-intensive retraining process observed in previous research [13,21]. This highlights the potential of the DER algorithm to enhance the classification of skin lesions and broaden the model’s ability to accurately identify various types of skin lesions. The results are significant for improving the efficiency and accuracy of skin cancer diagnoses, potentially aiding in early detection, timely treatment decisions, and better patient outcomes. The outcomes of this research could greatly influence the clinical setting by providing clinicians with a reliable and efficient tool for the automated classification of skin lesions. This could further facilitate early detection, timely diagnosis, and appropriate treatment decisions for patients with skin cancer.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by the Bio-Industrial Technology Development Program (No. 20014841) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

ORCID

Bong Kyung Jang (https://orcid.org/0000-0002-5723-4591)
Yu Rang Park (https://orcid.org/0000-0002-4210-2094)

References

4. Kumar P, Srivastava MM. Example mining for incremental learning in medical imaging. Proceedings of
2018 IEEE Symposium Series on Computational Intelligence (SSCI); 2018 Nov 18-21; Bangalore, India. p. 48-51. https://doi.org/10.1109/SSCI.2018.8628895


