Deep Skin Cancer Lesions Classification Scheme

Gabriel Chukwunonso Amaizu, Love Allen Chijioke Ahakonye, Dong-Seong Kim, and Jae-Min Lee
Department of IT Convergence Engineering
Kumoh National Institute of Technology
Email: (gabriel4amaizu, loveahakonye, dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—Skin cancer has risen to be one of the significant forms of cancer worldwide. However, the traditional means of skin cancer detection requires manual examination by an expert, which is often time-consuming. Recently, computer-aided methods have garnered much traction in the research community and among medical practitioners. This paper proposes using a deep-learning model to classify skin lesions into seven classes. The model is a fine-tuned 16-patch vision transformer (ViT) model and a custom classifier. The custom classifier consists of a convolutional computation fitted to the end of the ViT model. Experimental results show that the proposed scheme can classify various (7) skin lesions with an accuracy of 98%.**

***Index Terms*—classification, deep learning, skin cancer, vision transformer, CNN**

I. Introduction

Skin cancer accounts for approximately one-third of all reported cancer cases to the World Health Organization, and its incidence rate is increasing globally [1]. Countries like the USA, Australia, and Canada have experienced a noticeable rise in skin cancer cases over the last decade, with 15,000 people diagnosed each year, not surviving [2]. Specifically, melanoma cancer has been a significant concern, with 7,180 deaths reported in 2021 and projected to reach close to 7,650 deaths in 2022 [3], [4]. The loss of the ozone layer has led to an increase in dangerous ultraviolet (UV) radiation reaching the Earth's surface, posing a threat to skin cells and contributing to the development of cancerous cells [5]. Besides UV radiation, other factors such as smoking, alcohol use, infections, viruses, and the environment also influence the development of skin tumors.

Skin tumors can either be malignant or non-malignant. Among various types of malignant tumors, malignant melanoma, squamous cell cancer, dermatofibrosarcoma, and basal cell cancer are the most common [4]. Malignant melanoma, in particular, is considered the deadliest due to its potential for metastasis [4]. An early and accurate skin cancer diagnosis is crucial for successful treatment and improved patient outcomes, lowering mortality rates. Deep learning algorithms have shown promising results in medical imaging tasks, including skin cancer classification from dermoscopic images [4]. Nevertheless, their opaque nature hinders their adoption in clinical settings, making it challenging for medical professionals to comprehend and trust the model's predictions [6], [7].

Integrating artificial intelligence (AI) into clinical workflows has raised concerns about explainability [7], [8]. Deep learning algorithms have demonstrated higher accuracy in imaging applications than traditional machine learning methods. However, their lack of comprehensibility challenges clinicians, who may have trust issues with the classification model without clear explanations for their predictions [7], [9].

Vision Transformer (ViT) is a groundbreaking neural architecture that has revolutionized computer vision by applying the principles of self-attention mechanisms from natural language processing to image data. ViT breaks down an image into fixed-size patches, linearly embeds them, and then feeds them into a transformer encoder [10], [11]. This approach eliminates the need for hand-designed convolutional neural networks and enables end-to-end learning of image features. By facilitating bidirectional interactions between patches, ViT captures long-range dependencies, enabling superior modeling of global image context. ViT's success has pushed the boundaries of image classification and inspired advancements in various computer vision tasks, opening new avenues for AI research and applications [12]. This study uses a pretrained ViT model with a basic configuration and a custom classifier to detect skin lesions. Moreover, the custom classifier comprises a convolutional layer and a max pool layer, accompanied by a fully connected layer responsible for the classification process.

The study is thus arranged: Section I is preceded by a review of existing works on skin cancer classifications in Section II. Section III discusses the proposed methodology. Section IV highlights the experimentation environment and results. Section V concludes the study.

II. Related Works

In skin lesion classification, there has been a recent surge in the application of deep learning architectures and model interpretation. This section delves into some of the latest and most promising methods for effectively categorizing skin lesions and explainability of the model outcomes.

Leveraging a deep learning approach for automated skin cancer classification, [13] used deep convolutional neural networks to distinguish between benign and malignant cases with binary classification. The method achieved a notable test accuracy of 91.93% on the HAM10000 dataset. Another study by [14] presented a skin cancer classification methodology using deep learning on both HAM10000 and ISIC 2018 datasets.

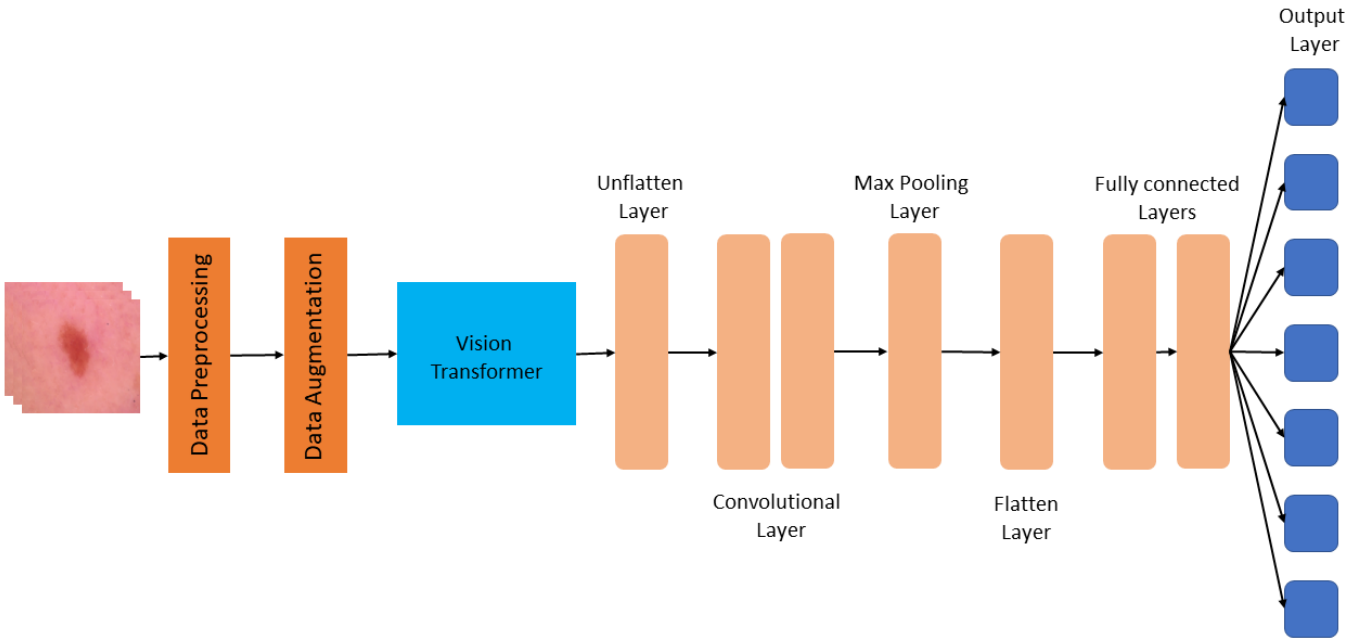

Fig. 1: Overview of the proposed scheme for skin lesions classification.

Feature extraction with a deep learning model and feature selection with a metaheuristic algorithm was employed, followed by extreme machine learning for the final classification, achieving accuracies of 93.40% and 94.36% for HAM10000 and ISIC 2018. Additionally, A multi-class skin classification approach based on deep learning models presented by [15] demonstrated improved accuracy through automated deep learning adjustments and ensemble models, with a remarkable accuracy of 93.20% on the HAM10000 dataset.

This study uses a pretrained ViT model with a basic configuration fitted with a custom classifier to detect skin lesions. Moreover, the custom classifier comprises a convolutional layer and a max pool layer, accompanied by a fully connected layer responsible for the classification process.

III. Methodology

In this section, we present a framework for accurately classifying skin lesions. Our proposed scheme combines deep-learning models to identify various skin lesions. The methodology outlined in this paper comprises a series of steps, as depicted in Figure 1. It begins with preprocessing and augmentation of the dataset. Next, we perform transfer learning by retraining a publicly available vision transformer model. This fine-tuned ViT model incorporates a convolutional layer, max pooling layer, and fully connected layer to address the classification problem. Ultimately, the model provides outputs for seven (7) distinct skin lesion classes. Further elaboration on these steps is provided in the subsequent subsections.

A. Dataset and Data Preprocessing

A dataset is a fundamental prerequisite for any deep learning model, forming the basis for training and validation processes. In this study, we employ the HAM10000 dataset [16] sourced from Kaggle. Kaggle, a distinguished platform within the scientific community, boasts an extensive repository of datasets spanning diverse domains, including the medical realm.

We engage in data augmentation to amplify the volume of training samples within the dataset. Concurrently, we deploy a Gabor filter to extract salient features within the images. The Gabor filter, a linear filter entrenched in image processing tasks, is brought into play. Through convolution with an input image, the Gabor filter engenders a filtered counterpart wherein specific frequency and orientation particulars undergo enhancement while concurrent suppression of alternative information transpires.

B. Modified Vision Transformer

Transformers, primarily rooted in self-attention mechanisms and initially harnessed for natural language processing, have found extensive utility in diverse computer vision tasks, consistently surpassing other deep neural networks like Convolutional Neural Networks (CNN) and Recurrent Neural Net-

works (RNN) [12]. A seminal concept underlying ViT models involves segmenting input images into fixed-size patches, subsequently subjecting these patches to a transformer-based architecture. This strategy facilitates extracting comprehensive global and intricate local features from the image [17].

In this investigation, we integrate a pre-trained vit_base_patch16_224 model, distinguished by its moderate layer count, attention heads, and associated parameters. With a patch size 16 and the capacity to process 224x224 images, this model has demonstrated proficiency on the ImageNet dataset. Thus, we undertake fine-tuning to tailor it for skin lesion classification. For this purpose, a custom classification head is introduced and linked to the terminus of the ViT model. This classification head comprises a series of discerning layers.

The initial layer unfurls the 1D array inherited from the preceding layer, yielding a 4D array, subsequently channeled into a convolutional layer. The convolutional output undergoes max pooling before flattening, followed by traversal through a fully connected layer encompassing two hidden layers. Ultimately, the output layer culminates in delineating one of seven distinct classes of skin lesions. This bespoke classifier effectively processes ViT model outputs, culminating in the ultimate prognostication of skin lesion types.

IV. Simulation and Performance Evaluation

As mentioned, the HAM10000 dataset, sourced from Kaggle, encompasses seven distinct classes. The dataset partitioning allocates 70% of the samples for training purposes, reserving the remaining 30% for validation. The computational experimentation occurred within a Jupyter environment, executed on an Ubuntu server that harnessed three NVIDIA GeForce RTX 3090 GPUs. During experimentation, the proposed model underwent 50 epochs of training, during which metrics such as accuracy, loss, and the confusion matrix were meticulously tracked. Table I encapsulates comprehensive insights into the model's statistics.

TABLE I: Model Statistics

Statistic	**Value**
Total Parameters	200,698,958
Trainable Parameters	200,698,958
Non-trainable Parameters	0
Input Size (MB)	0.57
Forward/Backward Pass Size (MB)	964.19
Parameters Size (MB)	765.61
Estimated Total Size (MB)	1730.37

In Fig. 2, the graphical representation illustrates the progressive trajectory of the model's training accuracy across 50 epochs. Notably, the model's accuracy commenced at an impressive 60%, exhibiting a steady ascent to approximately 95% by the 12th epoch. Subsequently, the accuracy reached its zenith at an impressive 98% at the culmination of the 50th epoch. In order to ensure efficient resource utilization, an early stopping mechanism was employed, effectively capping the model training at the 50-epoch mark after the point of negligible incremental accuracy gains.

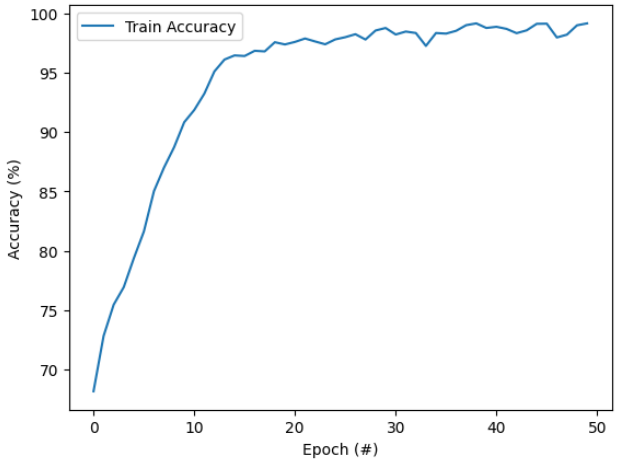

Fig. 2: Training accuracy after 50 epochs.

Minimizing loss is a core objective in deep learning. Fig. 3 presents the loss curve for the proposed framework. Initially, due to varying gradient selections, the loss starts high. However, by the 10th epoch, it decreases from 1 to 0.2; by the final epoch, it diminishes to less than 0.1.

The confusion matrix is a fundamental tool for assessing the performance of a classification model. It summarizes the model's predictions by comparing them to the true labels across different classes. The confusion matrix of the proposed scheme is in Fig. 4; each row of the matrix represents the instances in

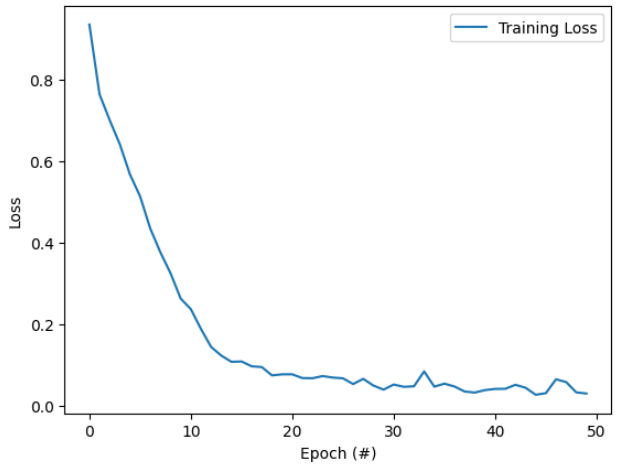

Fig. 3: Training loss after 50 epochs.

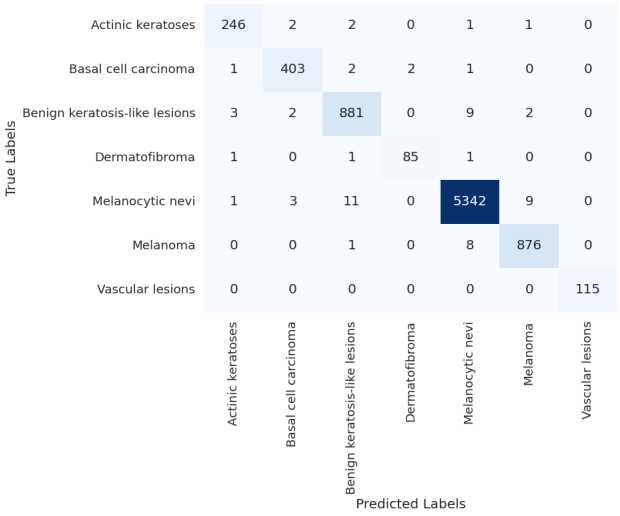

Fig. 4: Confusion Matrix.

a true class, while each column corresponds to the instances predicted by the model. The diagonal elements represent the true positive predictions, indicating the number of instances correctly classified for each class. Off-diagonal elements reveal misclassifications, where the row class was predicted as the column class.

V. Conclusion

This paper proposes a model for the detection of skin cancer. The model consists of a pretrained ViT model and a custom classifier. The ViT model is known to be robust and was initially trained using the ImageNet dataset. It has 16 patches and takes images of size 224X224. Moreover, the custom classifier consists of various layers, including convolutional, max pooling, and fully connected layers. The custom classifier is attached to the ViT model and is responsible for producing the final results of the skin lesions type. Simulation results indicated a high accuracy of the proposed model. The pretrained ViT fine-tuned using a custom classifier can be employed for multi-class skin lesions cancer detection.

Acknowledgement

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(MEST)(2018R1A6A1A03024003), the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation), and Industry-University-Research Institute platform cooperation R&D funded Korea Ministry of SMEs and Startups in 2022 (S3311338)

References

[1] S. A. AlSalman, T. M. Alkaff, T. Alzaid, and Y. Binamer, "Nonmelanoma Skin Cancer in Saudi Arabia: Single Center Experience," *Annals of Saudi Medicine*, vol. 38, no. 1, pp. 42–45, 2018.

[2] K. S. Nehal and C. K. Bichakjian, "Update on Keratinocyte Carcinomas," *New England Journal of Medicine*, vol. 379, no. 4, pp. 363–374, 2018.

[3] A. C. Society, "Key Statistics for Melanoma Skin Cancer," in *Facts & Figures*. American Cancer Society, Atlanta, Ga, January 12 2023.

[4] M. Zia Ur Rehman, F. Ahmed, S. A. Alsuhibany, S. S. Jamal, M. Zulfiqar Ali, and J. Ahmad, "Classification of Skin Cancer Lesions using Explainable Deep Learning," *Sensors*, vol. 22, no. 18, p. 6915, 2022.

[5] M. A. Albahar, "Skin Lesion Classification using Convolutional Neural Network with Novel Regularizer," *IEEE Access*, vol. 7, pp. 38 306–38 313, 2019.

[6] C. I. Nwakanma, L. A. C. Ahakonye, J. N. Njoku, J. C. Odirichukwu, S. A. Okolie, C. Uzondu, C. C. Ndubuisi Nweke, and D.-S. Kim, "Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review," *Applied Sciences*, vol. 13, no. 3, p. 1252, 2023.

[7] L. Zou, H. L. Goh, C. J. Y. Liew, J. L. Quah, G. T. Gu, J. J. Chew, M. P. Kumar, C. G. L. Ang, and A. W. A. Ta, "Ensemble Image Explainable AI (XAI) Algorithm for Severe Community-Acquired Pneumonia and COVID-19 Respiratory Infections," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 242–254, 2023.

[8] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What Do We Need to Build Explainable AI Systems for the Medical Domain?" *arXiv preprint arXiv:1712.09923*, 2017.

[9] K. Paranjape, M. Schinkel, and P. Nanayakkara, "Short Keynote Paper: Mainstreaming Personalized Healthcare–Transforming Healthcare Through New Era of Artificial Intelligence," *IEEE journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1860–1863, 2020.

[10] S. Qian, Y. Zhu, W. Li, M. Li, and J. Jia, "What Makes for Good Tokenizers in Vision Transformer?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–13, 2022.

[11] J. Ma, Y. Bai, B. Zhong, W. Zhang, T. Yao, and T. Mei, "Visualizing and Understanding Patch Interactions in Vision Transformer," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2023.

[12] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A Survey on Vision Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2023.

[13] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An Enhanced Technique of Skin Cancer Classification using Deep Convolutional Neural Network with Transfer Learning Models," *Machine Learning with Applications*, vol. 5, p. 100036, 2021.

[14] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H.-S. Yong, and J. Cha, "Multiclass Skin Lesion Classification using Hybrid Deep Features Selection and Extreme Learning Machine," *Sensors*, vol. 22, no. 3, p. 799, 2022.

[15] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A Multi-Class Skin Cancer Classification using Deep Convolutional Neural Networks," *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 28 477–28 498, 2020.

[16] 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.

[17] R. Garcia-Martin and R. Sanchez-Reillo, "Vision Transformers for Vein Biometric Recognition," *IEEE Access*, vol. 11, pp. 22 060–22 080, 2023.