DEEP LEARNING APPROACH FOR IMPROVEMENT IN MELANOMA DIAGNOSIS FOR DARK-SKINNED PATIENTS

Savannah Chan, Hana Kim, Hyun Seo Lee, Isalis Karhu-Leperd, Justin Wong
Johns Hopkins University
Machine Learning for Medical Application EN.520.439/659
{schan55, hkim350, hlee267, ikarhul1, twong53}@jh.edu

1 Motivation

Melanoma is diagnosed in approximately 100,000 patients each year in the United States, with lighter skin individuals at higher risk of developing the disease [1]. However, darker-skinned individuals are more likely to die from melanoma and frequently experience delayed detection due to serious diagnostic disparities [2] [3]. Current methods of diagnosing melanoma each have their own challenges: skin biopsy, the gold standard, is invasive and time-consuming, while non-invasive imaging techniques require expert interpretation [4]. AI-based approaches have been introduced to address these issues; however, because they are typically trained on images of lighter-skinned individuals, their diagnostic accuracy remains lower for darker-skinned patients [5] [6].

2 Literature Review

Previous research attempts to solve this disparity by developing new datasets, refining skin tone annotation methods, and implementing fairness metrics [7]. Barros et al. [6], for instance, curated a dataset of acral skin lesions specifically from black patients and evaluated both supervised and self-supervised deep learning models, which still revealed lower precision for darker skin tones. To expand data diversity, Rezk et al. [8] used VGG-19 trained on the ImageNet for dark-skin image generation, which was classified as benign or malignant skin lesions using a convolutional neural network (CNN). However, expert dermatologists noted that these synthetic images, while visually realistic, did not accurately capture disease presentation on darker skin. This gap in performance underscores the problem that AI models remain poorly generalized to underrepresented skin types due to limited and underrepresented training data.

3 Materials and Methodology

The main goal of our project is to ensure class balance in skin images through data augmentation and to use the generated images to improve generalization on melanoma classification. Before claiming that using deep learning methods would be the key to the whole solution, we explored classical data augmentation techniques, such as shifting colors, darkening, and adding Gaussian noise, as well as combinations of those techniques. However, these do not yield satisfactory data augmentation. As seen in Fig. 1), color shift has indeed changed the skin color by reducing redness and increasing yellowness. However, since our goal is to generate dark skin images, once we applied darkening or Gaussian noise, it has changed the skin color, but has made the boundary of the lesion stand out. Therefore, a machine learning-based approach is more suitable for solving this problem.

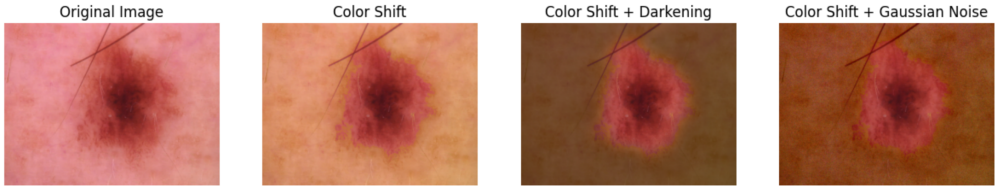

Figure 1: Classical Data Augmentation Techniques

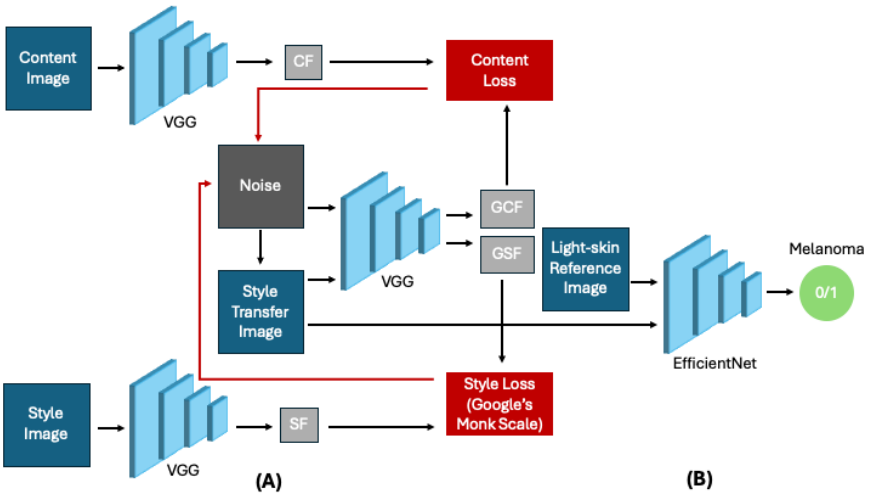

Figure 2: VGG-19 Model. (A) Architecture for feature extraction and style transfer. (B) Architecture for melanoma classification.

3.1 Model

To achieve realistic data augmentation, we will implement a data generation model based on style transfer using the pre-trained VGG-19 model [8]. This model will extract features from both the input image and a reference skin color image, ensuring that the input image (content) is transformed to have that target skin color (style). When computing the style loss, we will use the Google Monk Scale [9], which is a benchmark scale used to evaluate skin tones, to interpret the generated image's skin color.

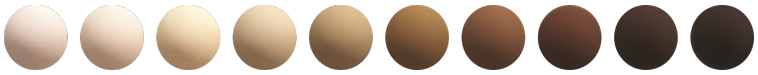

Figure 3: Google Monk Scale

3.2 Dataset

For our project, we will be using the HAM10000 dataset [10] for data generation, and CMB-MEL (Cancer Moonshot Biobank - Melanoma Collection) [11] along with other ISIC datasets [12] for the classification of melanoma.

To take a careful look at the data we're using for augmentation, we attempted to classify the images from the HAM10000 dataset against the Google Monk Scale [9]. The RGB values of the individual pixels in the scale images were extracted, allowing us to calculate the mean Euclidean distance between the RGB values of the 10 skin tones on the Monk scale and the input image. Originally, we used this method directly on the images from HAM10000, but we realised that it would take the pixels of the lesion into account as well, leading to inaccurate classifications of skin tone. We tested this out by classifying 2 images of the same skin lesion from the same patient, but were captured at a different time, and the two images were classified as different skin tones.

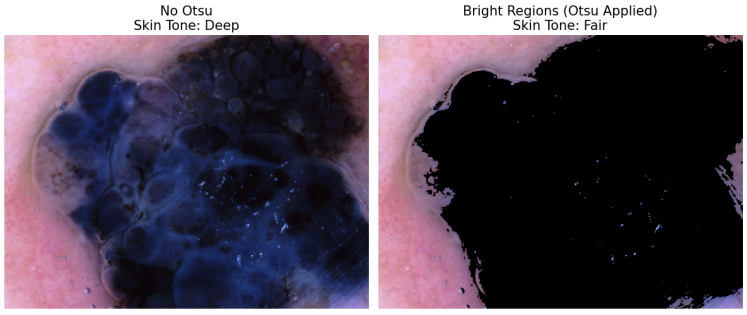

Figure 4: Skin colour classification with vs without Otsu correction

To resolve this error, we used Otsu thresholding to identify and remove the lesion before adjusting the skin tone. This ensures that the Euclidean distance is calculated based on the surrounding skin tone and not the lesion tone. As shown in Fig. 4, this method provides us with an accurate classification of the skin tone, regardless of the size of the lesion in the image.

Once we obtain the augmented dark skin images and achieve a balanced dataset, we will use them to train our melanoma classification model. The generated dark-skin images will only be used for training, while the data of real dark-skinned patients will be used for inference. Our classification model will be based on a pre-trained EfficientNet-B0 model, a Convolutional Neural Network (CNN) known for its high performance in classification tasks across various fields [13].

4 Significance

Through our results, our generated model aims to improve the accuracy of diagnosis of skin melanoma in dark-skinned patients, above the level of current methods. Our goal is to produce a generalizable model for melanoma classification. This would ultimately bridge the disparity in the rates of diagnosis and survival rates of melanoma patients between light-skinned patients and dark-skinned patients.

5 Preliminary Results

5.1 Dataset distribution for HAM10000 dataset

We analyzed the HAM10000 dataset distribution, in terms of age and gender, to determine whether our dataset is balanced. As seen in Fig. 5, age of the patients in the dataset is approximately normally distributed, with a peak in ages 40-50 years old and a lower representation of data in ages 0-20. This is due to the increased risk of melanoma in individuals as age increases, with the average age of melanoma diagnosis at 66 [14].

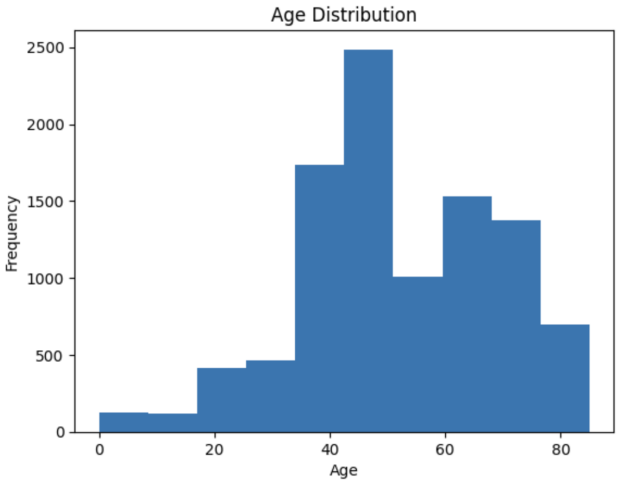

Figure 5: Histogram of age distribution

As shown in Fig. 6, the gender distribution is balanced, with 54% of males and 45.5% of females. Since melanoma exhibits different characteristics in men and women, ensuring balanced performance across gender is crucial. We will go on with our research to see if gender is indeed an important factor in melanoma classification.

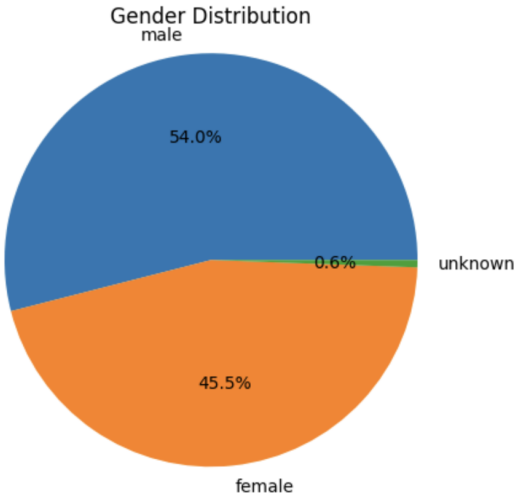

Figure 6: Pie chart of gender distribution

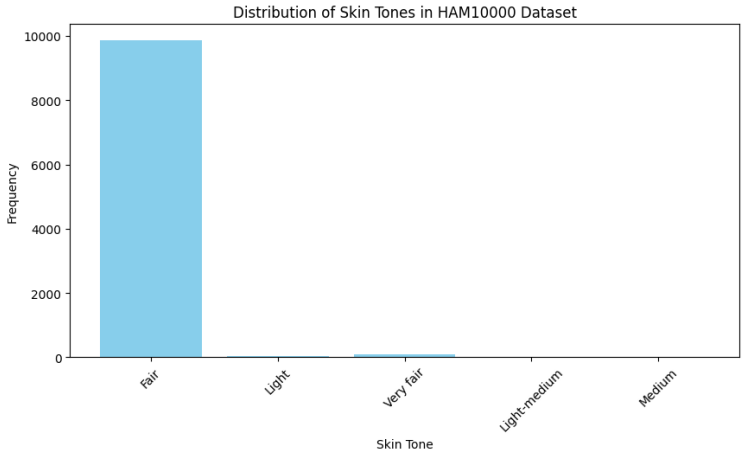

Figure 7: Distribution of skin tones

Furthermore, after running the images from the HAM10000 dataset against the Google Monk scale, we obtained a distribution of skin tones. As seen in Fig. 7, there is an overwhelming majority of fair skin tones (9,873 images) in the dataset, with almost no representation of darker skin tones. Specifically, there are only 48 light, 87 very fair, 6 light-medium, and just 1 medium skin tone image. This extreme imbalance in skin tone distribution further emphasizes the importance of generating more dark-skin tone data to create a more balanced dataset.

6 Next Steps

Due to the large majority of images in our datasets being fair skinned, we aim to utilize data augmentation through pre-trained VGG-19 models to generate dark-skinned images based on the original light-skinned images. Subsequently, we will combine the generated dark-skinned images with light-skinned images from other datasets to obtain a more balanced set of varying skin tones. Lastly, our goal is to generate a model trained on this new balanced dataset for classification of images as malignant or benign. We hope to create a more generalizable classification model which can be used on both dark and light-skinned patients.

References

[1] American Cancer Society. Key statistics for melanoma skin cancer. URL https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key-statistics.html. Accessed: 10 February 202.

[2] Gittleman H Barnholtz-Sloan JS Dawes SM, Tsai S and Bordeaux JS. Racial disparities in melanoma survival. *Journal of the American Academy of Dermatology*, 2016. doi: https://doi.org/10.1016/j.jaad.2016.06.006. URL https://www.sciencedirect.com/science/article/pii/S0190962216303802.

[3] Mehrotra Ravi Gupta Alpana, Bharadwaj Mausumi. Skin cancer concerns in people of color: Risk factors and prevention. *Asian Pacific journal of cancer prevention*, 2016. doi: https://doi.org/10.22034/APJCP.2016.17.12.5257. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5454668/#:~:text=Though%20people%20of%20color%20.

[4] Melanoma Research Alliance. Confirming the diagnosis, 2025. URL https://www.curemelanoma.org/patient-eng/diagnosing-melanoma/confirming-the-diagnosis. Accessed: 2025-02-10.

[5] Raj H. Patel, Emilie A. Foltz, Alexander Witkowski, and Joanna Ludzik. Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: A systematic review. *Cancers*, 2023. doi: 10.3390/cancers15194694.

[6] Luana Barros, Levy Chaves, and Sandra Avila. *Assessing the Generalizability of Deep Neural Networks-Based Models for Black Skin Lesions*, page 1–14. Springer Nature Switzerland, November 2023. ISBN 9783031492495. doi: 10.1007/978-3-031-49249-5_1. URL http://dx.doi.org/10.1007/978-3-031-49249-5_1.

[7] Laura N Montoya, Jennafer Shae Roberts, and Belen Sanchez Hidalgo. Towards fairness in ai for melanoma detection: Systemic review and recommendations, 2024. URL https://arxiv.org/abs/2411.12846.

[8] Eman Rezk, Mohamed Eltorki, and Wael El-Dakhakhni. Improving skin color diversity in cancer detection: Deep learning approach. *JMIR Dermatol*, 2022. URL https://derma.jmir.org/2022/3/e39143.

[9] Ellis Monk. Monk skin tone scale, 2019. URL https://skintone.google.

[10] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 2018. doi: 10.1038/sdata.2018.161. URL https://doi.org/10.1038/sdata.2018.161.

[11] Cancer Moonshot Biobank. Cancer moonshot biobank – melanoma collection (cmb-mel) (version 8) [dataset], 2022. URL https://doi.org/10.7937/GWSP-WH72.

[12] David Gutman, Noel C. F. Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint*, 2016.

[13] Rukhsar Sabir and Tahir Mehmood. Classification of melanoma skin cancer based on image data set using different neural networks. *Scientific Reports*, 2024. URL https://doi.org/10.1038/s41598-024-75143-4.

[14] Jeannette M. Olazagasti Lourido, Janice E. Ma, Christine M. Lohse, and Jerry D. Brewer. Increasing incidence of melanoma in the elderly: An epidemiological study in olmsted county, minnesota. *Mayo Clinic Proceedings*, 91(11):1555–1562, 2016. ISSN 0025-6196. doi: https://doi.org/10.1016/j.mayocp.2016.06.028. URL https://www.sciencedirect.com/science/article/pii/S0025619616303706.