Dermavision

Team Name: Salus

Team Members: Cindy Hua, Nina Dang, Mikayla Morrison
Mentors: Guy Morgenshtern, Atishya Reddy

Introduction

DermaVision addresses the critical shortage of dermatological care in Canada through a machine learning approach to skin condition diagnosis. This project combines image analysis with patient metadata to create a more accessible diagnostic tool for skin conditions. Our multi-input model integrates convolutional neural networks for image processing with metadata analysis, creating a comprehensive diagnostic tool.

The Issue

Skin diseases represent the 4th leading cause of non-fatal disease burden worldwide[1]. In Canada, fewer than 700 licensed dermatologists serve approximately 38 million people[2], resulting in wait times that extend beyond a year in major metropolitan areas[3]. This shortage disproportionately affects rural communities, where geographic barriers compound accessibility issues[4].

Proposal

DermaVision leverages multi-input machine learning to address dermatologist shortages and provide equitable care across diverse populations. The model consists of two branches: a Convolutional Neural Network (CNN) for processing image data, and a Multilayer Perceptron (MLP) for analyzing patient metadata. This method provides a multi-dimensional evaluation that can identify subtle patterns often missed by non-specialists. By flagging high-risk cases and offering preliminary assessments, DermaVision can promote early intervention, reduce wait times, and improve access to care in remote areas. Our tech stack includes: Python, TensorFlow/Keras, OpenCV, Scikit-learn, Pandas/NumPy, Matplotlib/Seaborn.

Data and Preprocessing

DermaVision utilizes the International Skin Imaging Collaboration (ISIC) dataset, the largest public collection of skin lesion images, comprising 503,955 images accompanied by clinical metadata. Each image's metadata includes 30 features, 20 of which were deemed relevant for data preprocessing. Using a Decision Tree, five features were identified as essential for model development: Sex, Age, Fitzpatrick Skin Type, Anatomical Site, and Lesion Size. Although the dataset originally contained over 500,000 records, the model required only 600 features for training. Ultimately, 626 records included all five essential features and were used to train the model.

A key challenge with the metadata was the lack of correlation among features. A confusion matrix revealed that most features exhibited weak positive or negative correlations. This suggested that the final model would be more complex than initially anticipated. Additionally, the metadata demonstrated limited diversity in Fitzpatrick skin type. Although the scale typically includes six skin tone categories, only four

---

[1] Seth, D., Cheldize, K., Brown, D., & Freeman, E. E. (2017, August 7). *Global burden of skin disease: Inequities and innovations - current dermatology reports*. SpringerLink. https://link.springer.com/article/10.1007/s13671-017-0192-7

[2] Canadian Medical Association. (2019). *CMA*. Canadian Medical Association. https://www.cma.ca/sites/default/files/2019-11/2019-01-spec-prov_1.pdf

[3] SkinMD. (2024, August 31). *Canada's dermatologist shortage and Crisis*. DermCafé. https://www.dermcafe.ca/post/canada-dermatologist-shortage

[4] Marchesan, J. (2024, May 25). *Ontario facing growing dermatologist shortage*. CityNews Toronto. https://toronto.citynews.ca/2024/05/25/ontario-facing-growing-dermatologist-shortage/

were represented in the dataset. Medium tones were sparsely included, and deep tones were absent. This lack of representation proved to be a critical limitation, as the imbalance in skin tone distribution led to fewer training instances for darker skin types, resulting in poor model performance on these tones in the test set.

Methodologies

The mixed-input model combines a CNN branch and an MLP branch to process skin lesion images and clinical metadata simultaneously. Features from both branches are concatenated before passing through the final classification layers.

The CNN branch utilizes MobileNetV2 pre-trained on ImageNet with frozen weights for transfer learning. Input images (224×224×3) are processed through the base model, followed by global average pooling for flattening before passing through the dense layer.

The MLP branch processes six clinical features (anatomical site, dermoscopic type, skin type, sex, age, lesion size) through two fully-connected hidden layers. Each layer employs ReLU activation, batch normalization, and dropout to prevent overfitting.

Results

We tested three variations: one without image segmentation optimizing for accuracy, another with segmentation maximizing recall (to minimize false negatives or predicting benign when it is malignant), and a third with segmentation optimizing for accuracy.

The three models show a start of good learning curve convergence, though we would need more epochs to say for sure. The non-image segmentation approach got the highest test accuracy at 85%.
There is a bit of a gap between training and validation accuracy, indicating some overfitting. It is also visible in the loss curve where there is lower training loss compared to the unseen data, affecting its ability to generalize well. The loss curve for the image segmentation maximizing recall model appears to have validation loss plateaus, which again suggests the model may be overfitting.

The dataset imbalance likely contributed to the models' weaker performance on malignant classification. This imbalance is reflected in the F1-scores, which consistently showed a substantial performance gap between benign (0.90-0.91) and malignant (0.52-0.64) classification across all model variations,

Conclusion

Future endeavours for this project include refining the model and enhancing its ability to identify specific categories of melanoma. Most importantly, it is essential to improve the model's transparency, allowing patients and healthcare practitioners to understand how it arrives at each conclusion.