A Web Application for Detecting Skin Disease in Women of Colour

Fabiha Fairuzz Subha [1], *Ebunoluwa Makinde*[2]

Abstract

Since the advancement of artificial intelligence (AI), there has been a rise in skin diagnostic tools curated to assist in dermatological care. However, the datasets often used to create these technologies highly underrepresent certain skin types. Specifically, skin tones that represent people of colour. In our work, we introduce a proof of concept web-based application that uses deep learning for skin disease classification specifically for women of colour. We achieve this by using novel and diverse datasets alongside curating a dataset of synthetic images that represent these more underrepresented skin tones.

Introduction

Skin diseases impact billions worldwide, yet many lack access to adequate dermatological care [3]. The World Health Organization predicts a rise in conditions like melanoma, expected to cause over 100,000 deaths annually by 2040 [12]. Even in developed countries, long wait times and limited resources highlight the need for accessible diagnostic tools. While AI has shown promise in dermatology, existing tools are trained on datasets with limited diversity, primarily featuring lighter skin tones. This results in reduced diagnostic accuracy for individuals with darker skin tones.

Our project aims to address this disparity by developing a web-based AI diagnostic tool to detect skin diseases in individuals with diverse skin tones, focusing on people of color. Using Flask for the backend and React for the frontend, the tool will incorporate inclusive datasets and advanced convolutional neural networks (CNNs). With a user-friendly interface, the application seeks to ensure equitable access to AI-driven dermatological care globally.

Related Work

Improving Skin Tone Diversity

Rezk et al. [11] addressed the issue of the underrepresentation of darker skin images in diagnosing skin pathology in people of colour by generating images for these underrepresented skin tones using artificial intelligence. They propose four phases for their methodology: identifying the skin tones underrepresented in current data, generating images for these skin tones, evaluating the quality of the generated images, and developing a classification model using both synthetic and original data. They conclude their work by noting that phase 1 is complete, phase 2 is in progress, with phases 3 and 4 to be completed in their future work.

In phase 1, they segment the data by separating the disease region from the skin to analyze the skin colour. A clustering method is then applied to categorize the data into groups, followed by developing a classification model to determine the skin tone type. For phase 2, they propose using style transfer and deep blending image generation methods, both based on feature extraction and utilizing the Visual Geometry Group (VGG) network.

Generating Synthetic Images

Patcharapimpisut and Khanarsa utilize the HAM10000 dataset, which consists of seven skin diseases, to train a stable diffusion model for synthesizing images related to these conditions[8]. They extend their work by training two classification models using InceptionResNetV2 and ResNet50. Their results show that the synthetic images they generated when combined with real images, improved the model's recall and accuracy.

A skin lesion style-based generative adversarial network (GAN) was developed by Qin et al. for generating high-quality skin lesion images[9]. Their network was trained using the International Skin Imaging Collaboration (ISIC) 2018 classification challenge[2]. They then utilized transfer learning with ResNet50 for a classification task and found that the performance improved when augmenting with the synthetic dataset they curated.

[1]Undergraduate student, Department of Computer Science, University of Calgary, fabihafairuzz.subha@ucalgary.ca
[2]Undergraduate student, Department of Computer Science, University of Calgary, ebunoluwa.makinde@ucalgary.ca

Applications for Skin Disease Detection

Hussain et al. developed a mobile application for classifying skin diseases using a convolutional neural network (CNN)[4]. Their architecture includes three convolutional layers, max pooling layers, and an output layer for classification. This model was trained on the Ham10000 dataset and the ImageNet dataset. They then integrated the trained model into a mobile app built with React Native using TensorFlow Lite. The app allows users to capture images from the gallery or camera; after which, the model predicts the results, displaying information about symptoms and precautions.

Methodology

Data Acquisition and Preprocessing

The dataset we currently utilize is the Skin Condition Image Network (SCIN) dataset from Google, which contains 10,408 crowd-sourced dermatological images[13]. SCIN is considered to be more representative and diverse compared to existing datasets in the literature and has a predominantly female contributor base (66.72%). Thus, it effectively represents the demographic we target with our application.

The SCIN dataset is public and contains the metadata and labels of the dataset, which includes the path to the images, their labels, and other information such as skin type based on the Fitzpatrick skin type (FST) classification. Each image is diagnosed by 1 to 3 dermatologists; at least one label is assigned to the images with a confidence score between 1 and 5. The data is provided in a dataframe and the column weighted_skin_condition_label, indicates the combined differential diagnoses by the dermatologists which are normalized to sum to 1. This fictitious example elaborates more on the SCIN schema [13]:

- dermatologist_skin_condition_label_name: [Nummular eczema, Acute and chronic dermatitis, Psoriasis vulgaris, Eczematous dermatitis, Post-inflammatory hyperpigmentation]
- dermatologist_skin_condition_confidence: [3, 2, 1, 3, 4, 3]
- weighted_skin_condition_label: {'Eczema': 0.69, 'Acute and chronic dermatitis': 0.11, 'Post-inflammatory hyperpigmentation': 0.11, 'Psoriasis': 0.08}

As we explored the SCIN dataset, we encountered issues stemming from its significant class imbalance, which led to suboptimal performance during our initial model training. Consequently, for our final model iterations, we augmented our dataset with the dataset curated by [5]. Their collection comprises a total of 1,657 images representing six types of skin diseases: eczema, acne, rosacea, keratosis, milia, and carcinoma. Additionally, one class is designated for images that do not depict any of these diseases. These images were sourced from public dermatologist datasets as well as self-collected resources [5].

To combine both datasets, we sifted through the SCIN dataset, focusing specifically on images with a maximum differential diagnosis weight of 0.5 or greater. Accordingly, we labeled the images based on this criterion, ensuring that the condition matched one of the six present in the second dataset. We also selected images featuring Fitzpatrick skin types 4, 5, or 6 to capture the darker skin tones not represented in the second dataset. The final dataset comprised 1,773 images, with 1,241 allocated for training, 266 for validation, and 266 for testing. A custom PyTorch dataset class was created to access the data while training our deep learning model. This dataset class is capable of splitting the data into training, test, and validation sets. Lastly, we applied transformations to resize each image to 224 by 224 pixels and normalized each image before passing it to the model for training.

However, only 32.6% of the contributors to SCIN are of non-white racial identity. To address this lack of representation, we attempt to generate datasets that include darker skin colors, making our dataset more robust for these underrepresented racial groups. For generating our synthetic data, we narrowed our approach to generate images for the cases that were reported as having eczema in the differential diagnosis. To prepare the dataset for training our GAN, we filtered the SCIN dataset to focus on images with eczema in the differential analysis. In order to enhance the dataset for darker skin tones, we also filtered based on the self-reported FST6 skin type, which is the darkest on the FST scale. This resulted in 319 images that we used to train our GAN.

Deep Learning Models

To determine the diagnosis of skin diseases, we implement transfer learning using the pretrained network provided by torchvision, ConvNext Small [7], utilizing weights trained on the IMAGENET V1 dataset [10].

In our transfer learning approach, we freeze all layers except for the classifier layer to preserve the information from the pretrained network. We modified the classifier to match the custom classifier created by [5]. Since they successfully used it with the TensorFlow network XCeption to classify the same skin diseases we aim to classify, we adopted a similar approach [1, 5]. The final output layer is adjusted to correspond to the number of classes (diseases) in our dataset, which consists of 6 diseases plus an additional class for instances where no disease is identified. For all models, we employ the Adam optimizer with a learning rate of 0.001. After training, we conduct a validation run for each epoch and, upon obtaining our best model, we further evaluate its performance by testing on a dataset that the model has not encountered, thereby assessing its performance on new data.

For data generation, we employ a GAN instead of other synthetic data methods like Stable Diffusion, primarily due to its lower computational cost. We utilize StyleGAN2-ADA in PyTorch [6]. The model is trained continuously to ensure proper GAN training, taking a snapshot of the generated images after every 10 iterations or epochs to evaluate the quality of the generated data. We also save the model at this stage, allowing for continued training if the runtime is interrupted.

Web Application for Skin Disease Detection

To provide a user-friendly interface for our skin disease detection tool, we developed a web application using Flask for the backend and React for the frontend. This application serves as a bridge between users and the trained deep learning models, enabling them to upload dermatological images and receive diagnostic predictions seamlessly.

Backend Development

The backend for this project is implemented using Flask, a lightweight Python web framework, and integrates machine learning models for image-based disease classification. The system enables communication with a frontend client and processes user-uploaded image files to predict skin conditions. Key components and functionality are outlined below:

- **Framework and Libraries:**
  - **Flask:** Provides the web server and API endpoints for interacting with the model.
  - **Flask-CORS:** Ensures Cross-Origin Resource Sharing to allow secure communication between the backend and frontend.
  - **Pillow:** Facilitates image file handling and preprocessing.
  - **PyTorch:** Powers the machine learning model for efficient prediction and inference.
- **Machine Learning Model**

  A custom deep learning model is implemented using PyTorch, built upon the pre-trained ConvNeXt Small architecture, with the following design:
  - **Base Model:** Pre-trained on ImageNet, with frozen parameters to utilize its feature extraction capabilities.
  - **Custom Classifier:** Composed of fully connected layers, batch normalization, dropout layers, and softmax activation to classify skin conditions into seven categories: Acne, Carcinoma, Eczema, Keratosis, Milia, None, and Rosacea.
  - **Trained Model:** The trained model, saved as *mekha_7.0.pt*, is loaded at runtime for inference.
- **Preprocessing and Transformation**

  Uploaded images undergo the following transformations to align with the input requirements of the model:
  1. Resizing to 224×224 pixels.
  2. Conversion to a tensor.
  3. Normalization using ImageNet mean and standard deviation values.
- **API Endpoint**

  /predict Endpoint

– Accepts an image file as input via a POST request.
– Preprocesses the image and performs inference using the trained model.
– Returns a JSON response containing the predicted class.

- **Class Mapping**

  The model output indices are mapped to human-readable class names using a predefined dictionary.

Frontend Development

The React-based frontend provides an intuitive and user-friendly interface, enabling seamless interaction with the application. Users can easily upload images of their skin conditions through the web interface, which are then sent to the backend API for processing. Once the backend processes the image, the frontend displays the predicted skin condition label prominently, offering immediate and actionable feedback. The application features an **Upload Image Page** that allows users to upload an image through a drag-and-drop or file selection option, ensuring a smooth and responsive experience. After uploading, the image is passed to the trained CNN model for classification, and the results are displayed clearly on the same page.

In addition to the upload functionality, the application includes an Informative Page called **Common Diseases** that provides users with detailed explanations of the potential causes and symptoms associated with the predicted skin condition. This feature goes beyond mere predictions, offering educational insights to enhance user understanding and awareness. It also includes helpful guidance on when to seek medical advice, empowering users to make informed decisions about their health. The frontend design is clean, modern, and fully responsive, ensuring accessibility across various devices, including desktops, tablets, and smartphones. Progress indicators provide real-time feedback during the image upload and processing stages, while error messages inform users of any issues, such as invalid image uploads.

Experimental Analysis

Results

The final model curated was assigned a classification task, where each image given to the model was labeled with one of six common skin disease types or categorized as ”none” if no identifiable skin disease was found. With our implemented custom model, we successfully trained it using 200 epochs, enabling it to classify the images according to the diseases. The best model achieved an accuracy of 0.891, a recall of 0.887, precision of 0.889, and an F1 score of 0.887. The total model training took about 8 hours on a L4 GPU on Google Colab.

Regarding our GAN for generating synthetic images, we conducted over 30 hours of training, completing 320 iterations and processing a total of 1,280,000 images (1,280kimg). To quantitatively assess performance, we employed the Fréchet Inception Distance (FID) metric to measure the closeness of generated data to real data. A lower FID indicates that the generated synthetic images are closer to the real images. For our GAN training, we achieved a minimum FID score of 115.76. We display some real and fake images generated by our model in Fig X, showcasing the visual similarities between the generated images and the real ones.

Comparative Analysis

Our underlying research goal was not only to develop a model for predicting skin diseases but also to create one that can overcome the biases present in the current data used to train AI models. This would enable it to identify diseases across a broader spectrum of skin tones. To achieve this, we aimed to compare the effectiveness of our classifier model in identifying diseases in darker skin tones, which are often underrepresented in publicly available datasets. To conduct this comparison, we trained a second model using the same architecture as our original model and ran it through 250 epochs. This model was trained on a dataset that did not include the augmented SCIN images. We then curated a small test dataset consisting of 10 images of darker skin tones depicting acne and eczema—none of which the models had encountered during training—and passed these images through both models to obtain their predictions.

As illustrated in Figure 2, it is clear that our model outperformed the baseline model trained without the augmented SCIN dataset in predicting skin diseases in darker skin tones. Our model was 3 times more accurate in predicting the presence of eczema and 3.5 times more accurate in identifying acne in these skin

Figure 1: Real Images from the SCIN dataset and fake images generated by our GAN using random seeds

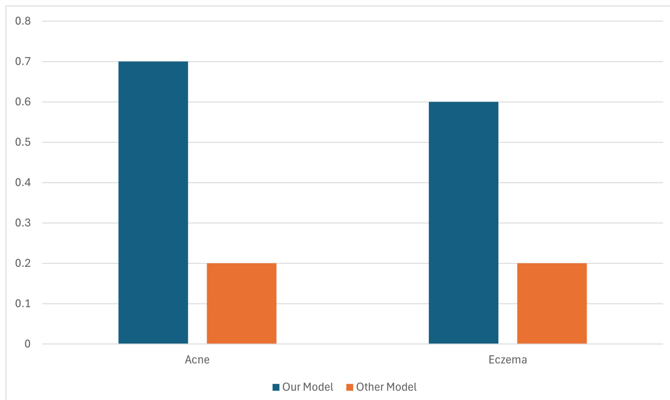

Figure 2: Comparison of the prediction accuracy of the presence of Acne and Eczema on darker skin tones for our model trained on the SCIN augmented dataset and the model trained on the dataset without SCIN

tones. This highlights the bias inherent in existing datasets and how the model's accuracy can be skewed when trained on data that does not represent the diverse global population.

Contributions and Future Implementation

In this study, we have made the following contributions to the field of dermatology:

1. Curated an augmented dataset combining public dermatology data with a more novel dataset, SCIN, to ensure representation of diverse skin tones.
2. Successfully trained a classifier model to predict six classes of skin diseases based on our enhanced dataset.
3. Developed a GAN model to demonstrate proof of concept for generating images of darker skin tones that are commonly underrepresented in dermatology datasets.
4. Created a functional and dynamic web application that enables users to upload images and utilize our model for predicting skin diseases.

Overall, we have surpassed previous models documented in the literature by successfully detecting skin diseases across a broader range of skin types, as evidenced by our comparative analysis results.

However, while our classification model is functional, several challenges and limitations remain. First, our model is limited to the number of diseases it can identify, which could be expanded by providing a more robust dataset encompassing a greater variety of conditions. Although our model shows improved accuracy in detecting skin diseases on darker skin tones, it is not flawless, and training it on a dataset that more accurately reflects a wider range of skin tones would be ideal. Additionally, our GAN, designed to diversify the dataset, struggled to generate sufficiently realistic images suitable for training. Increasing the size of the training datasets and extending training durations could enhance its performance. Exploring alternative image generation techniques, such as diffusion models, may also yield improvements in quality.

Conclusion

This project demonstrates the potential of leveraging artificial intelligence to address healthcare disparities, specifically in dermatology. By integrating diverse datasets and employing advanced deep learning techniques, the web application effectively detects skin diseases in individuals with underrepresented skin tones. The implementation of a convolutional neural network (CNN) built on the ConvNeXt Small architecture showcases the feasibility of transfer learning for accurate disease classification. The addition of a generative adversarial network (GAN) to augment the dataset further strengthens the model's capacity to address biases inherent in existing data.

The user-centric design of the web application ensures that individuals can seamlessly upload images, receive predictions, and access informative content about their skin condition. This accessibility empowers users to make informed health decisions while raising awareness about dermatological issues. The project's focus on diverse skin tones sets it apart from existing tools, addressing an essential gap in AI-driven dermatological care.

While the results are promising, challenges such as dataset limitations and the generation of realistic synthetic images remain. Future work could include expanding the dataset to cover a broader range of diseases, enhancing the GAN's performance, and exploring alternative data augmentation techniques. By continuing to refine these aspects, this project can contribute significantly to the equitable application of AI in healthcare and set a precedent for developing inclusive diagnostic tools.

References

[1] François Chollet. "Xception: Deep learning with depthwise separable convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.

[2] Noel Codella et al. "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)". In: *arXiv preprint arXiv:1902.03368* (2019).

[3] Roxana Daneshjou et al. "Disparities in dermatology AI performance on a diverse, curated clinical image set". In: *Science advances* 8.31 (2022), eabq6147.

[4] Muhammad Zunnurain Hussain et al. "Mobile application using cnn for skin disease classification with user privacy". In: *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*. IEEE. 2023, pp. 52–57.

[5] Nadav Ishai et al. "Teledermatologist: Automated Diagnosis of Skin Diseases with Image Recognition". In: *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE. 2023, pp. 1–1.

[6] Tero Karras et al. "Training Generative Adversarial Networks with Limited Data". In: *Proc. NeurIPS*. 2020.

[7] Zhuang Liu et al. "A convnet for the 2020s". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 11976–11986.

[8] Parapat Patcharapimpisut and Paisit Khanarsa. "Generating Synthetic Images Using Stable Diffusion Model for Skin Lesion Classification". In: *2024 16th International Conference on Knowledge and Smart Technology (KST)*. IEEE. 2024, pp. 184–189.

[9] Zhiwei Qin et al. "A GAN-based image synthesis method for skin lesion classification". In: *Computer Methods and Programs in Biomedicine* 195 (2020), p. 105568.

[10] Benjamin Recht et al. *Do ImageNet Classifiers Generalize to ImageNet?* 2019. arXiv: `1902.10811 [cs.CV]`. URL: `https://arxiv.org/abs/1902.10811`.

[11] Eman Rezk, Mohamed Eltorki, Wael El-Dakhakhni, et al. "Improving skin color diversity in cancer detection: deep learning approach". In: *JMIR Dermatology* 5.3 (2022), e39143.

[12] Jie Sun et al. "Machine learning methods in skin disease recognition: a systematic review". In: *Processes* 11.4 (2023), p. 1003.

[13] Abbi Ward et al. *Crowdsourcing Dermatology Images with Google Search Ads: Creating a Real-World Skin Condition Dataset*. 2024. arXiv: `2402.18545 [cs.CY]`. URL: `https://arxiv.org/abs/2402.18545`.