Addressing Bias and Data Scarcity in AI-Based Skin Disease Diagnosis with Non-Dermoscopic Images

Chiara Bellatreccia, Andrea Borghesi, Roberta Calegari

Alma Mater Studiorum
University of Bologna

March 19, 2025

Overview

Introduction - The Problem and The Goal

The Problem

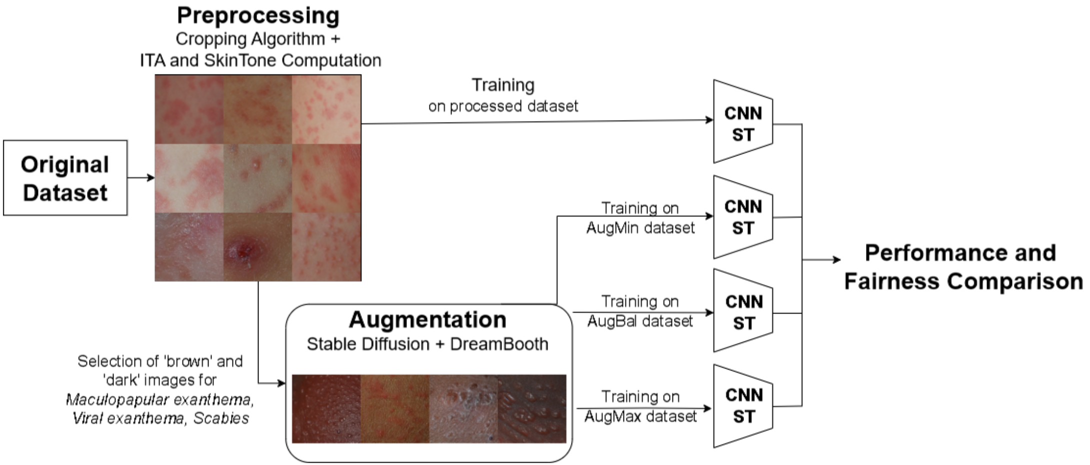
The imbalance in the skin tone attribute in dermatological datasets leads to possibly biased classification

The Goal

Augment a non-dermoscopic imbalanced dataset using synthetic images generated by fine-tuning a Stable Diffusion model with DreamBooth and compare the classification bias between the augmented and original dataset using:

- A **Convolutional Neural Network (CNN)**
- A **Swin Transformer (ST)**

Introduction - The Pipeline

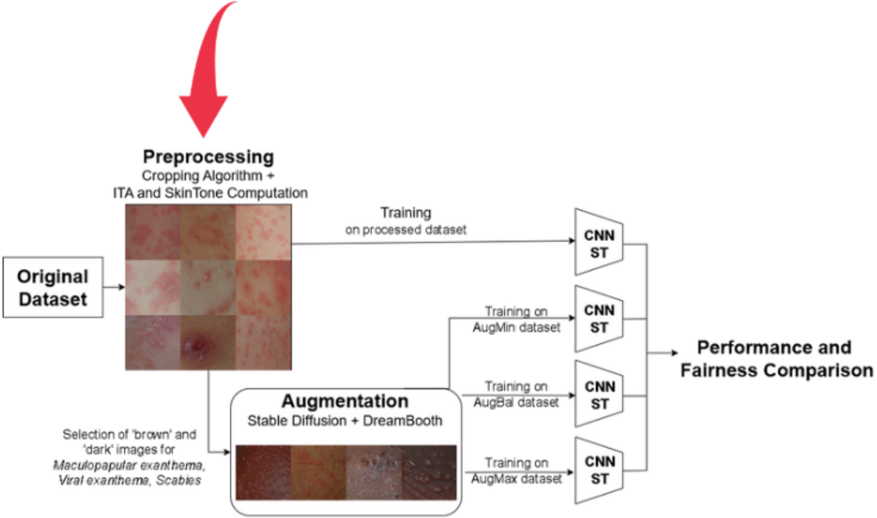


Introduction - The Dataset

Non-dermoscopic dataset of $\sim$ 8000 images of pediatric patients at the Sant'Orsola Hospital with nine possible skin diseases.

- High variability in illumination
- High variability in size and quality
- Inconsistent focus on affected skin areas
- Blurriness
- Imbalance in skin tone
- Imbalance in disease classes

Data Preprocessing: Cropping Algorithm and Skin Tone Estimation



Data Preprocessing: Cropping Algorithm and Skin Tone Estimation

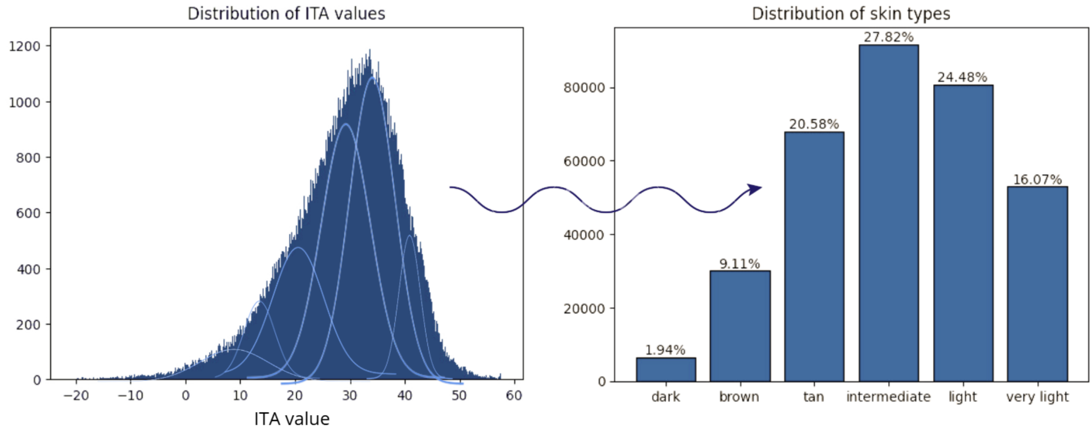
Sliding window-based algorithm[1] which employs the **binary mask** of the images to detect the regions in which the disease is present

- **RESULTS:** $\sim$ 330000 crops of uniform 256x256 size. Not all crops are good crops:

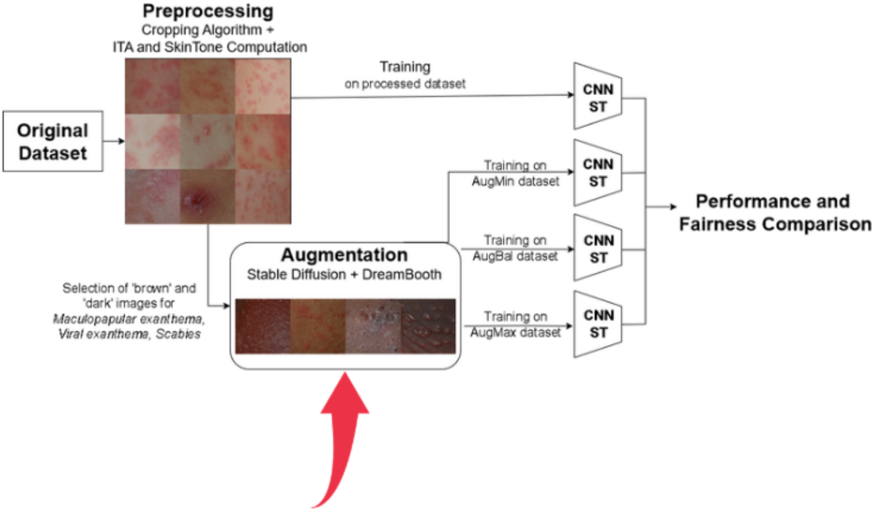
[1]The preprocessing algorithm was adapted from the one developed by D'Amico, Murgia and Moeini Feizabadi

Data Preprocessing: Cropping Algorithm and Skin Tone Estimation

Individual Typology Angle computation and thresholding using a Gaussian Mixture Model to obtain the skin tone labels



Generation of Clinical Skin Images: Fine-Tuning Stable Diffusion

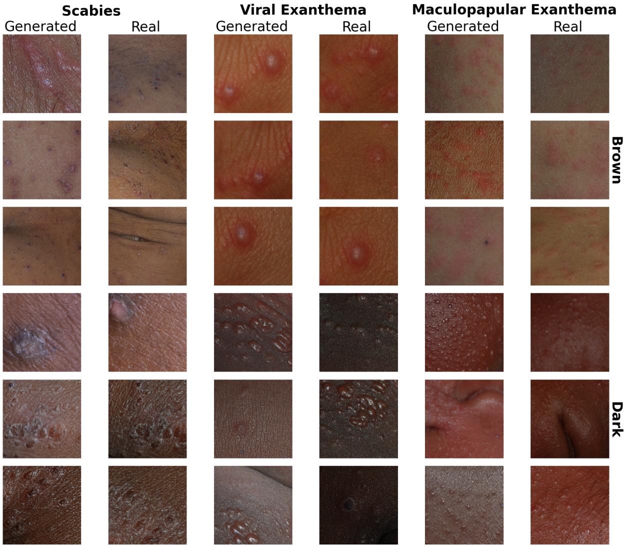


Generation of Clinical Skin Images: Fine-Tuning Stable Diffusion

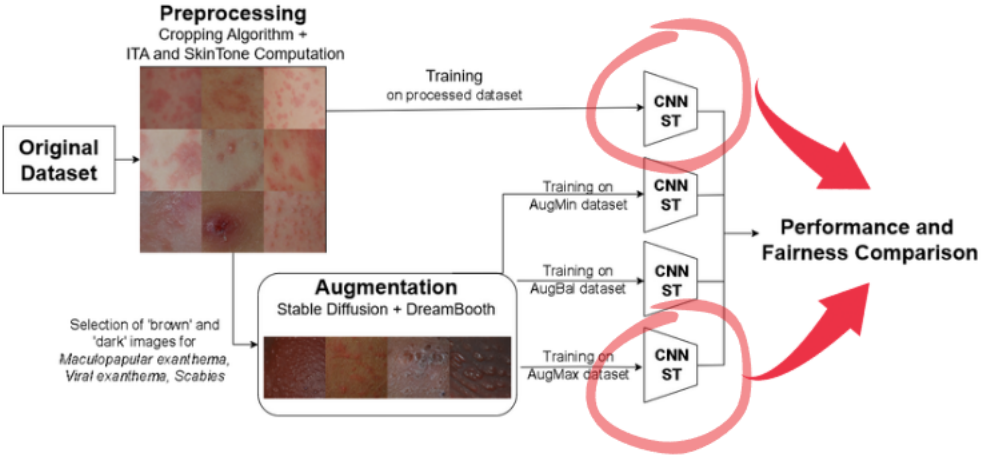
Fine-tuning of Stable Diffusion via the DreamBooth technique:

1. Exploration of the dataset to **manually** select images of 'dark' and 'brown' skin tones
2. Construction of mini-datasets ($\sim$ 14-29 images) for each disease and each of the two skin tones
3. **Grid Search** during DreamBooth fine-tuning to find the set of optimal hyperparameters
4. **Manual** selection of the models fine-tuned with the hyperparameter combinations showing the best results

Generation of Clinical Skin Images: Fine-Tuning Stable Diffusion



Best CNN Results



Skin Diseases Classification with a CNN: Original vs Augmented Dataset

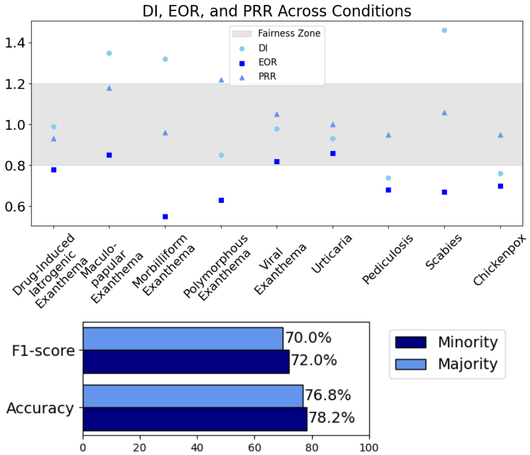
Three fairness metrics to evaluate bias:

1. Disparate Impact
2. Equalized Odds Ratio
3. Predictive Rate Ratio

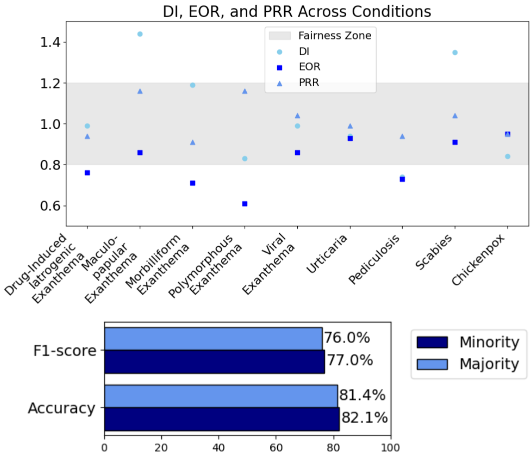
Subdivision of the skin tones in a **Minority** group (*dark* and *brown* skin tones) and a **Majority** group (*tan*, *intermediate*, *light* and *very light* skin tones)[2]

[2] A. Corbin, O. Marques, Assessing bias in skin lesion classifiers with contemporary deep learning and post-hoc explainability techniques, IEEE Access 11 (2023) 78339–78352. doi:10.1109/ACCESS.2023.3289320.

Skin Diseases Classification with a CNN: Original vs Augmented Dataset



(a) CNN results on the **original dataset**.

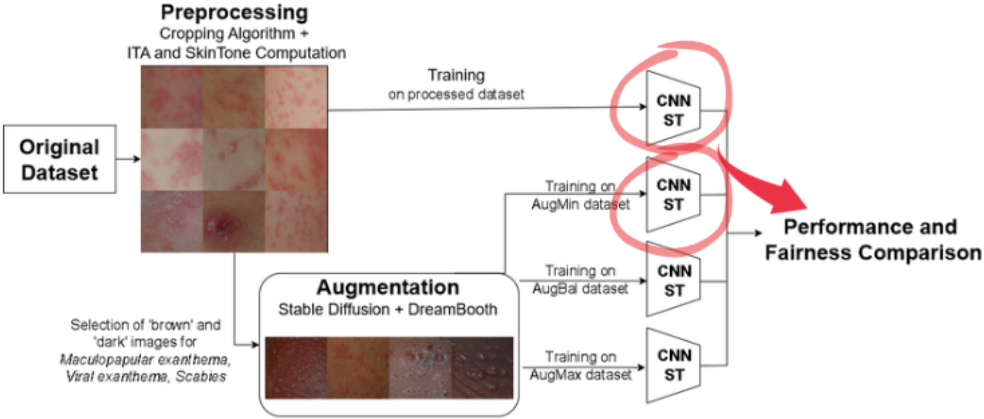


(b) CNN results on the **augmented dataset**.

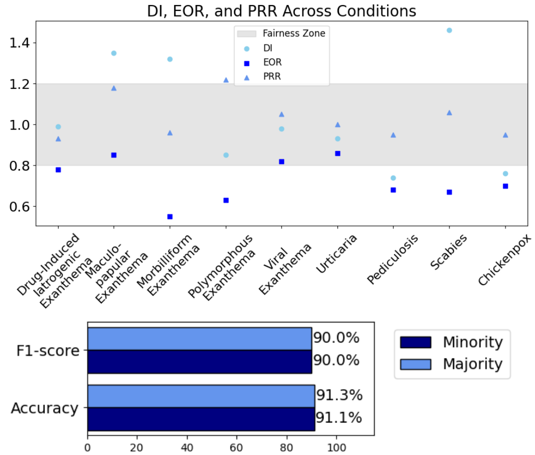
17 out of 27 fairness metrics values improved

Skin Diseases Classification with a ST: Original vs Augmented Dataset

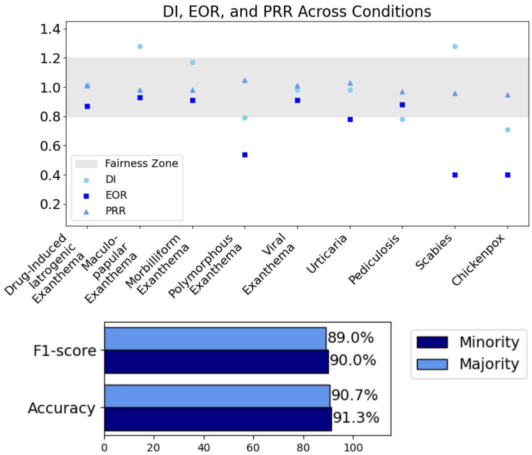
Best ST Results



Skin Diseases Classification with a ST: Original vs Augmented Dataset



(a) ST results on the **original dataset**.



(b) ST results on the **augmented dataset**.

15 out of 27 fairness metrics values improved

Conclusions and Further Work

The results demonstrate that synthetic images have a positive impact on classification, both in terms of **performance** and **fairness**. The improvement is more remarkable for the CNN than for the ST. However, several improvements to the pipeline are possible:

- Improve preprocessing to enhance classifier performance
- Generate dark-skinned images starting from light-skinned images... but pay attention!

Thank you for your attention!