Preliminary Assignment

Projects in Data Science

28. February 2026

Technical details:

LearnIt: Damian Wojciech Wardzinski, GitHub: wardziun, Annotation ID: A
LearnIt: Hunor Szabo, GitHub: Hounczy, Annotation ID: B
LearnIt: Maksymilian Rafal Wozniak, GitHub: makswozniak, Annotation ID: C
LearnIt: Marcus Bülow Pedersen, GitHub: marcuswhg295 , Annotation ID: D
LearnIt: Simon Friland, GitHub: sifriland, Annotation ID: E

Repository can be found here: `https://github.com/Hounczy/2026-PDS-Group_M`. Our annotations can be found on the main branch of our GitHub repository, with the name of "combined_annotations.csv". Within the different folders there are few files. These are just placeholder files with no meaningful contribution to our preliminary assignment.

Group's combined summary

Introduction

As part of our data science project, we have been provided with medical images containing various skin lesions. The purpose of this report is to give a brief overview of the dataset's content and to detail our primary objective: annotating the images for the presence of hair and pen markings.

Data Overview

Our group (M) was assigned a specific subset containing 117 images of skin lesions. The photos are taken with a phone to capture the texture, color, and borders of the affected skin.

Data Quality and Image Characteristics

Image sharpness varies considerably across the dataset. While some images (e.g., PAT_2144_4738_120) are exceptionally clear with high contrast, allowing for easy identification of the lesion border and surrounding skin, others suffer from noticeable blur. For instance, images like PAT_2116_4614_671 and PAT_2013_4148_303 appear slightly out of focus. This blurriness can complicate the process of distinguishing the skin lesions from general skin texture. Lightning is inconsistent throughout the dataset. Some images are taken with good lightning, while others appear underexposed or taken in low-light environments. Furthermore the pictures have also been taken from different distances, angles and phone cameras. This can make the lesions look larger or smaller, and can have an effect on the quality of the data.

Observations and Conclusion

Successfully annotating the 117 images required careful attention to detail. Identifying pen markings was generally straightforward, as the blue or purple ink strongly contrasts with human skin. However, grading hair on a 0 to 3 scale proved more subjective and required a standardized approach in order to distinguish barely visible hair on the over-lit pictures. We believe that pen markings could increase the quality of data by excluding the skin lesions from the surroundings. But there is also random pen marks that could proof to be more harmful than useful. By accurately annotating our groups pictures, the dataset can be used to train models more effectively.