Group Q report

Krzysztof Marcinkiewicz (krzysztt0)
Dumitru Russu (Dumitru-dev-44)
Harini Kothi (Harini128)
Marius Grossu (mgrossu)
Laurits Richter (LauritsRich)
https://github.com/OpenDataScienceGroup/2026-PDS-Q.git

1 Image–Mask Alignment Review

Images 1–20: Most masks aligned well with their corresponding lesions. However, in image *104-1755-320.png*, the mask was entirely black, meaning the lesion was not identified. For image *138-205-296.png*, the mask failed to capture the main lesion and instead highlighted eight separate, distinct areas.

Images 21–40: Most masks corresponded accurately to their images. However, several were inaccurate due to interference from pen marks or hair occlusion, which affected proper lesion identification.

Images 41–60: The majority of masks fit the lesions well, though some imperfections were observed. In lesion 717, the mask did not capture the affected area at the top edge of the image. In lesion 3924, only the most prominent lesion was identified, while other similar areas were overlooked.

Images 81–100: Although most masks aligned well with the lesions, this section was heavily affected by clinical artifacts such as purple pen marks and thick hair. These artifacts may create "shortcuts," leading the model to associate ink or hair with malignancy rather than focusing on the lesion itself.

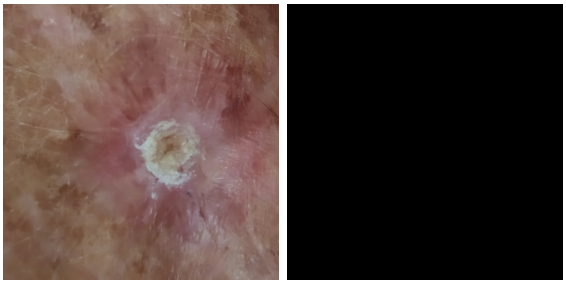

Figure 1: Failed mask

2 Insights from Metadata

The dataset primarily represents an elderly population, with most patients between 60 and 80 years old and a slight male majority (57.5%). Most individuals reported no history of smoking or drinking; however, approximately 31% of the smoking and drinking data is missing.

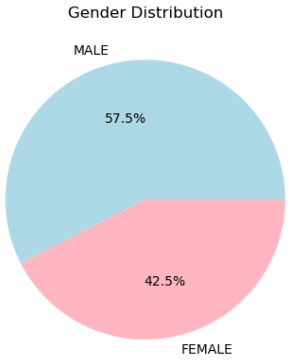

Figure 2: Gender distribution

The age distribution further confirms that the dataset is dominated by older individuals.

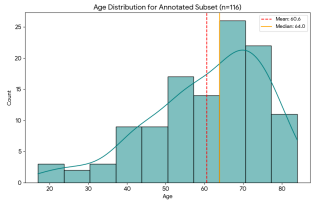

Figure 3: Age distribution

The most common diagnoses are **BCC** and **ACK**, both of which are strongly associated with long-term sun exposure in older adults. Overall, the dataset reflects a high-risk, aging population, consistent with the high prevalence of BCC and ACK cases observed.

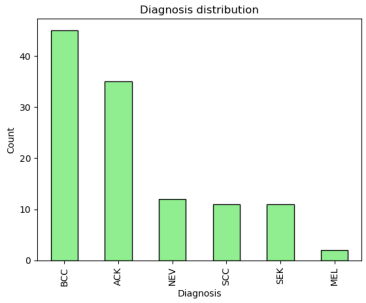

Figure 4: Diagnosis distribution

3 Insights from Annotations

3.1 Hair Coverage Levels

Most images were annotated as having low hair coverage, which is beneficial for model training. However, images classified as hair coverage levels 2–3 may require further consideration regarding inclusion or preprocessing.

3.2 Pen Marks

Approximately one-fifth of the images contain pen marks, which could negatively affect model performance if not addressed.

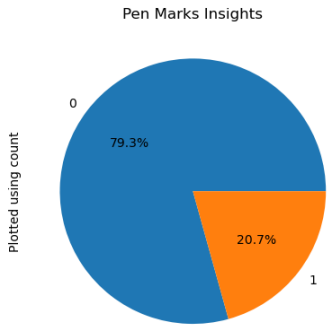

Figure 5: Pen marks distribution

3.3 Inter-Annotator Agreement

Inter-annotator agreement, measured using Cohen's kappa, showed:

- **Substantial agreement** for hair coverage ($\kappa = 0.6997$)

- **Almost perfect agreement** for pen mark detection ($\kappa = 0.8613$)

Hair coverage assessment is inherently more subjective, whereas pen mark identification is more straightforward. This likely explains the higher agreement observed for pen mark detection.