Problem Statement Worksheet (Hypothesis Formation)

Can we create a deep learning model that can classify images of skin lesions into "malignant" and "benign" categories, with associated likelihoods?

1 Context

Skin cancer is the most prevalent type of cancer. Melanoma in particular is responsible for 75% of all skin cancer deaths, despite being the least common type of skin cancer. If caught early, most Melanomas can be cured with minor surgery. Leveraging AI to aid in early and accurate diagnosis of melanoma has the potential to save lives. (source: SIIM and ISIC)

2 Criteria for success

We will maximize the AUROC between the predicted probability and the observed target.
Because we're interested in minimizing false negatives in this medical diagnostic problem, we will also use recall as a KPI
Ideally, we'll also push this model onto a hosted website to create a webapp people can try out with their own images

3 Scope of solution space

We will focus primarily on the binary target of "benign" or "malignant"

4 Constraints within solution space

- Our model has to be better at classifying malignant lesions then the average dermatologist to be useful in a medical setting
- We may not be able to generalize to smartphone images using dermoscopic images

5 Stakeholders to provide key insight

My Springboard Mentor
A great example project: https://github.com/Tirth27/Skin-Cancer-Classification-using-Deep-Learning
Fast.AI course and book: https://course.fast.ai/
https://github.com/fastai/fastbook
Nature paper published from Stanford: Dermatologist-level classification of skin cancer with deep neural networks

6 Key data sources

Kaggle Comp: SIIM-ISIC Melanoma Classification

33,100 training image files, available as .jpg with .csv metadata, .DICOM, OR .tfrecf.

Original ISIC source

The ISIC 2020 Challenge Dataset https://doi.org/10.34970/2020-ds01 (c) by ISDIS, 2020