Capstone 3 Project Documentation

Problem Statement

The most common cancer in the United States is skin cancer, with a prevalence of 20% by the age of 70 [1]. Early detection is essential for prompt treatment and survival. A model to predict malignant skin cancer from at-home images can accelerate the process of seeking medical support.

Approach & Results

Machine learning model was built using 2 datasets and ~10,000 images using a Convolutional Neural Network with a focus on optimizing recall for malignant cases. The resulting model has a recall of 0.96.

Recommendations for Use

Incorporation with a mobile app for household implementation.
The model performs best on patients with lighter skin tones due to training data.

Recommendations for Further Research

Augmentation of training data with cancer images on darker skinned individuals. To note: skin cancer is harder to detect on darker skinned individuals and data may thus be limited [2].
Additional model for skin cancer that develops under the nail.

Data Sources

- Malignant vs benign skin cancer images: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign
- Additional dataset: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

References

1. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/
2. https://www.skincancer.org/skin-cancer-information/skin-cancer-skin-of-color/