



Skin Cancer Detection

Capstone 3
Golden Se





Introduction/Background



Introduction/Background

- Most common cancer in the US: skin cancer
 - 20% chance by the age of 70
- Early detection essential for prompt treatment and survival



Proposed Solution

Proposed Solution

- A model to predict malignant skin cancer from at-home images can accelerate the process of seeking medical support.





Data Sources



Data Sources

- The International Skin Imaging Collaboration archive
- Harvard HAM10000 dataset of skin lesions

ISIC_0033619.jpg





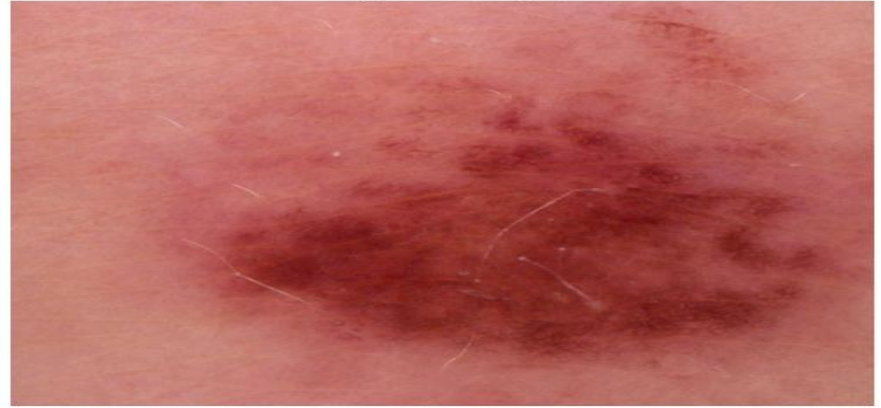
Data Processing & Insights



Data Processing & Insight

- Images resized, normalized
- HAM dataset split into “malignant” or “benign” to match other dataset
- Mild class imbalance: 25.92% malignant
- Images are of lighter skin tone

ISIC_0025558.jpg



ISIC_0027423.jpg





Model Building & Evaluation



Model Building & Evaluation

- Convolutional neural network
- Loss function: Binary Cross Entropy with Logits Loss
- Optimizer: SGD (Stochastic Gradient Descent)



Results & Recommendations for Future Work



Results

	precision	recall	f1-score	support
0	0.97	0.40	0.56	1960
1	0.36	0.97	0.53	702
accuracy			0.55	2662
macro avg	0.67	0.68	0.55	2662
weighted avg	0.81	0.55	0.55	2662

Recommendations for Future Work

- Incorporation with a mobile app for household implementation
- Model performs best on patients with lighter skin tones
- Recommendations:
 - 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
 - Additional model for cancer that develops under the nail

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/>
3. <https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign>
4. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>