Ryan Slattery

Cancer Detection Capstone Final Report

Problem Statement:

"Skin cancer is a major public health problem, with over 5,000,000 newly diagnosed cases in the United States every year. Melanoma is the deadliest form of skin cancer, responsible for an overwhelming majority of skin cancer deaths. In 2015, the global incidence of melanoma was estimated to be over 350,000 cases, with almost 60,000 deaths. Although the mortality is significant, when detected early, melanoma survival exceeds 95%." (ISIC 2018 | ISIC 2018: Skin Lesion Analysis towards Melanoma Detection)

To maximize the chance of skin cancer being detected early, the barrier to testing should be as low as possible. To help enable this, ISIC released HAM10000, a dataset of over 10000 labeled skin lesion images belonging to 7 classes, 3 of which are cancerous and 4 of which are benign. This dataset could be used to train a deep learning-based image classifier. Such a classifier could be made available for use through a smartphone app to make screening highly accessible.

Data Cleaning and Preparation:

As a first step, I verified that all images were accounted for and all had labels. Next, I had to make my train/validation/test split while accounting for the fact that my dataset contains multiple images of the same lesion, ensuring we don't get images of the same lesion in multiple subsets of the data. Then I created weighted data loaders so that classes would appear with equal frequency during training. These data loaders also apply transformations to the data when the data is loaded during training. These transformations include normalization as well as random mirroring and rotation.

Models:

I started the modeling process by training a simple baseline using a LeNet like architecture.

Next I trained a ResNet-based model. The model used the pretrained ResNet-50 from the official PyTorch model zoo. All layers were frozen except for bottleneck layers 3 and 4 as well as the final output layer.

More technical details about these can be found in model_metrics.txt

Findings:

The LeNet like architecture had a validation accuracy of 49%. The ResNet had a validation accuracy of 74% and a test accuracy of 76%.

Recommendations:

1) Don't use this as a replacement for an exam conducted by a physician.
2) Deploy in a system that doesn't suffer excessively from the low accuracy. For example: Display an output of "Could not be identified" if we can't confidently rule out cancer and "See a doctor" if the model outputs cancer as a prediction.
3) Find a way to improve model performance by collecting more data or improving data efficiency.

Further Research:

My final accuracy leaves a lot to be desired in terms of both overall accuracy and the recall of cancerous classes. This is in part due to model overfitting that I encountered during training. Using a smaller model didn't improve results here but more data might. Unfortunately, the available data is limited. Fortunately, we might we able to produce synthetic data that could help with training by training a generative adversarial network using one of the methods overviewed here:
https://medium.com/abacus-ai/gans-for-data-augmentation-21a69de6c60b

Another way to address this might be to reframe the problem as a binary classification problem where the two classes are cancerous and benign.