Detecting Cancer with CNNs - Final Report

Author: Jared Lauer

1.0 Problem Statement

The goal of this project is to evaluate the use of Convolutional Neural Networks (CNNs) to reduce the workload of medical pathologists who screen lymph node histopathology samples under a microscope to search for the presence of metastatic cancer in breast cancer patients. When performed manually, this process can be time-consuming and prone to error, and with recent advances in slide scanning technology and the cost reduction of digital storage in recent years, automating this process has become increasingly feasible.[1]

2.0 Background

According to the Canadian Cancer Society, 1 in 8 Canadian women will develop breast cancer in their lifetime, and 1 in 33 Canadian women will die from it.[2] When breast cancer advances from stage 1 to stage 2, it begins to metastasize, and the first place this is detectable is in the closest lymph node to the tumour. The 5 year survival rate drops from 98% to 88% when the cancer enters stage 2, and by stage 4 where the cancer has spread to other major organs, the survival rate drops to 16%.[3] It is critical to identify stage 2 cancer at the earliest possible opportunity to provide the best outcomes for the patient.

In this project, I am using the Patch Camelyon (PCam) dataset which was made publicly available by Bastiaan Veeling et al. from Cornell University.[4] This dataset contains 327, 680 colour images extracted from 400 whole slide images of stained histopathology samples extracted from the lymph nodes of breast cancer patients. Each image is 96 x 96 pixels in RGB colour, and the dataset is about 7.48 GB in size. The images are labelled as Class 0, containing healthy tissue, and Class 1, containing at least 1 pixel of cancer tissue within the central 32 x 32 pixel region of the image. There are multiple sources to obtain this data, but in my project I imported it through TensorFlow's dataset catalog. The dataset is split into train, validation, and test sets with a split of approximately 80%/10%/10%. This dataset did not require much preprocessing or Exploratory Data Analysis (EDA), but I visualized samples of each split and confirmed that the class balance for each set is almost exactly 50/50.

CNNs are a popular choice for image classification tasks, and tend to outperform many other machine learning models because they preserve the spatial relationships which exist in image data while leveraging the predictive power of neural networks.

---

[1] https://jamanetwork.com/journals/jama/fullarticle/2665774
[2] https://www.cancer.ca/en/cancer-information/cancer-type/breast/statistics/?region=on
[3] https://www.sciencephoto.com/media/1137060/view/breast-cancer-stages-illustration
[4] https://arxiv.org/abs/1806.03962

3.0 Modelling Summary

I evaluated 3 different CNN architectures on this dataset using Google Colab. The first is CNN1 which I optimized based off of a simplified VGG16-like architecture proposed by Geert Litjens[5] and trained entirely on the PCam dataset. The others are the deep neural networks VGG16 and ResNet50 which were trained on the ImageNet dataset and adapted for PCam with transfer learning techniques. The PCam dataset has been benchmarked by a team of 11 pathologists who were given roughly 2 hours to classify the slides in the test set, and they were evaluated based on their recall (sensitivity) of the positive class, and the area under their respective receiver operating characteristic curves (AUC). A summary of these results follows below.

	Pathologist	CNN1	VGG16	ResNet50
AUC	0.810 (mean)	0.906	0.900	0.872
Recall	62.8% (mean)	93.5%	71.2%	68.4%

Figure 1. Results Summary

All three CNN architectures outperformed the pathologist benchmark, but CNN1 performed the best by a wide margin. This is likely due to the fact that the ImageNet weights were not particularly effective at extracting features from the PCam dataset, as the ImageNet images do not closely resemble histopathology slides. The two transfer learning networks also proved extremely prone to overfitting on the training set, and it may be possible to better optimize their performance in the future.

4.0 Conclusion

CNNs present a viable alternative to the manual process by which pathologists screen histopathology samples. However, due to the severe consequences of missing cancer cells in these samples, it is important to prioritize the recall of the models on the positive class above all else. These networks would be best utilized as a pre-screening tool that reduces a pathologist's workload by rejecting healthy samples, with a pathologist manually reviewing these samples afterwards.

There are a number of ways to expand this project in the future. I would like to test these CNNs on grayscale images, as the colour of the images is irrelevant to the class labelling. I would also like to analyze only the central 32 x 32 pixel region of the images to see if this improves performance by eliminating unnecessary information. Finally, I would like to identify commonly misclassified images across all of the models, and visualize the convolutional filters of these networks to gain additional insights into how they are making their predictions.

[5] https://geertlitjens.nl/post/getting-started-with-camelyon/