BreastNet: A CNN Model for Breast Cancer Prognosis

Abstract

Cancer that is detected in its early stages can give patients more options and save thousands of dollars in medical costs. Some of the most recent developments in computer science and machine learning are in the biomedical field, especially individualized healthcare. There is also an increase in the demand for telehealth options, reducing healthcare costs. With the help of computational technology, medical practitioners will be able to process data more quickly, which will allow more patients to have access to reliable treatment. Besides, systematic processes for interpreting various data types (such as clinical features, genetic information, and medical images) can identify trends that a human eye would not detect.

Introduction

When patients are diagnosed with breast cancer, it is important for them to understand how serious the disease is, and they rely on experts like doctors and radiologists to interpret a wide range of medical data in their record, which includes high quality images as well as clinical data and patient history. By understanding the risk score of high and low risk, the patients are able to make the right decisions about their future treatment plans. CNNs have the ability to make use of medical images from a variety of sources to aid in the diagnosis and prognosis of diseases like cancer. Since breast cancer is so common in women, there are large publicly available datasets that contain a large number of Magnetic Resonance Images (MRI). The I-SPY clinical trial was used to train BreastNet, and it is a publicly available dataset which is available through The Cancer Imaging Archive.

Research Question(s)

How well can we predict the 5 year life expectancy of Breast cancer based on patients’ MRI and clinical data?

We used MRI and clinical data from publicly available datasets found through The Cancer Imaging Archive. The ACRIN I-SPY clinical trial included MRI images and survival information that was used to model a binary classifier based on censored survival data. Data was processed into training and test directories, each containing two survival classifications.

Materials and Methods

The initial datasets were composed of folders containing slices of DICOM data that were organized by patient, time relative to treatment, and view type (eg. Scout, Sagittal, etc.). Since the model was constructed using Keras and TensorFlow libraries, we needed a dataset that was in a standardized format for processing. We created a preprocessing program using the pydicom library to convert a DICOM directory from a series of 2D Sagittal view images to a 3D numpay array. We also processed a separate “region of interest” (ROI) image, which was provided by radiologists, into a 3D numpay array. The ROI could be used as a bitmask, which can be used to focus the training of the model on image details directly related to the breast cancer tumors, which can reduce some of the noise found in the high-dimensional data. Some of the benefits of using 3D numpay arrays include the fact that it can be sliced in an number of dimensions, and that it can be easily processed by python scripts, which can open the problem of medical image analysis to include computer scientists who use python as their primary tools of choice.

Model

We created a prototype for a model that took in 3D MRI images containing different scans of breast cancer. The model is a shallow 3D Convolutional Neural Network that accepts numpay array inputs with the shape $(1 \times 256 \times 256 \times 36)$.

Data Preprocessing

We created a prototype for a model that took in 3D MRI images containing different scans of breast cancer. The model is a shallow 3D Convolutional Neural Network that accepts numpay array inputs with the shape $(1 \times 256 \times 256 \times 36)$.

Challenges

Some of the major challenges related to this project include the fact that Keras does not natively support 3D image preprocessing. Some of these steps must be done as a result, we needed to use custom libraries to generate data to train the network.

Contact Information

Ryan Deem's email: ryan.deem@students.kennesaw.edu
Cora Meador email: cora.meador@students.kennesaw.edu

Acknowledgments

We would like to thank our advisor Mohammed Aledhari for his support for our work, and ths KSU College of Computer Science and Software Engineering

Conclusions

BreastNet contributes to a growing body of works that use machine learning to help cancer patients understand their diagnosis and make informed treatment decisions. By understanding whether they have a high or low risk of death within the next five years, they will have the agency to maximize their quality of life. By focusing on specific results that patients find most important when learning about their breast cancer diagnosis, BreastNet can help make a positive impact on people’s lives; one that goes beyond the simple results the program generates.

References

[10] "The Cancer Imaging Archive. The ACRIN I-SPY clinical trial included MRI images and survival information that was used to model a binary classifier based on censored survival data. Data was processed into training and test directories, each containing two survival classifications.

Authors: Ryan Deem and Cora Meador
Advisor: Dr. Mohammed Aledhari