

CS-6440: Sprint 1 – Practicum Project Proposal

Joseph Waugh

jwaugh6@gatech.edu

1 AREAS OF FOCUS

The proposed area of focus is health applications in machine learning, and cloud computing environments. The primary topic within this area is dealing with machine learning and artificial intelligence within health applications. The specific area of focus here is creating a deep learning model aimed at predicting the presence of colon cancer in early stages via CT colonography scans. I am looking to approach this problem as an individual project.

2 BACKGROUND AND SIGNIFICANCE

The usage of CT colonography scans is a 21st century solution for screening patients for colorectal neoplasms. In a recent study concluded in 2009, CT colonography tests among 49 individual case studies proved extremely effective, where “no cancers were missed at CT colonography (scans)” (Pickhardt, et.al.), versus outdated technologies that have proven less effective.¹ While this new technology has proven itself to be very promising, colorectal cancer remains one of the worst types of cancer in terms of mortality rates in the western world.² Prior methods involved fecal tests, which have proven effective in terms of reducing mortality rates for individuals diagnosed with colon cancer; however, due to the increased sensitivity of the CT colonography scans, there’s a better possibility for reducing mortality rates, and the technology has proven to be accurate in recent trials, both of which suggest that the new technology should be the path forward in terms of screening.

¹ Colorectal Cancer: CT Colonography and Colonoscopy for Detection—Systematic Review and Meta-Analysis; Perry J. Pickhardt, Cesare Hassan, Steve Halligan, and Riccardo Marmo; Radiology 2011 259:2, 393-405

² Ries, L. A. G. (2000, May 15). The annual report to the nation on the status of cancer, 1973â€“1997, with a special section on colorectal cancer. American Cancer Society Journals. <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.1002/%28SICI%291097-0142%2820000515%2988%3A10%3C2398%3A%3AAID-CNCR26%3E3.0.CO%3B2-I>

3 PROBLEM

One of the issues with the current, manual process of reviewing any sort of medical scans, including CT colonography scans, is the presence of false positives and true negatives. These errors are both costly for the patient and provider and minimizing these while still utilizing a manual process will never work as best as it is intended; however, this proves another point, that accuracy in the interpretation of these CT scans needs to be of high importance as well. This leaves two primary ideas to solve for: creating a timely, yet accurate solution to identify colon cancer early, by encouraging frequent scans and providing an easy method to complete this action.

4 PROPOSED SOLUTION OR IDEA

To address the issues related to ease of use, encouraging frequent CT colonoscopy testing, and providing accurate results and feedback, the proposed solution is to implement a recurrent neural network model via a web application that allows a user to upload images of their CT scans, and the model will output a binary decision, whether the user is showing signs of colorectal cancer, or if they appear to be healthy of this specific type of cancer, with the contingency being that a particular confidence interval needs to be shared (i.e. these results are 85% accurate, etc.). Depending on the result, this information can be handled two different ways: if the user's test result appears healthy, they're given the option to send themselves a reminder to complete the test again within the next 12 months based on a user-specified date, in addition to listing warning signs to look for in the meantime; however, if the test indicates the possible presence of cancer, then the user is prompted to send this information to their personal doctor, in addition to retaking the test as another option, and having information to the nearest medical center if necessary. This will mean that there's multiple pages within the web application, with the model being hosted on the backend, with a trained dataset of these CT scan images from the Cancer Imaging Archive, which contains multiple thousands of photos for the model to train (and test) on.

5 COMPLEXITY OR EFFORT

In terms of project complexity, to implement the solution, a machine learning model will be implemented, and hosted online via a flask-based web application that is run through HTML and Python scripts. A potential complexity with this solution is hosting the dataset that the model is trained on, or finding a way to host the trained neural network model without the dataset, will require some effort to determine. Building the different web pages for the application as well will create some complexity, given the model will be hosted on the main page, and the result of the model will determine how the website will react. This is an area of complexity, and when configuring the page layout for actions required by the user if a test indicates cancer may be present, finding out how this may work based on in-network versus outside-network providers, based on a user's insurance type. This may be outside the scope of an individual project, but it will be scoped for the initial stage of the project. When the user is submitting their data to run through the model, another concern from the user is data privacy, and determining how their data will be used and if it will be linked to them in any way, is a major factor that needs to be addressed. Solutions aimed at removing any personal information from the scan will be another area of emphasis, and this will be another feature this is explained within the web application itself to remove any concerns related to the model's processes.

6 TENTATIVE TEAM MEMBERS & ROLES

I plan on working on this project individually. Therefore, I will be handling all associated responsibilities.

7 REFERENCES

1. Colorectal Cancer: CT Colonography and Colonoscopy for Detection—Systematic Review and Meta-Analysis; Perry J. Pickhardt, Cesare Hassan, Steve Halligan, and Riccardo Marmo; Radiology 2011 259:2, 393-405
2. Ries, L. A. G. (2000, May 15). The annual report to the nation on the status of cancer, 1997, with a special section on colorectal cancer. American Cancer Society Journals.