Cancer AI Summary

Silba Dowell, Milana Wolff, Daniel Hintz, Shad Sellers, Nile Young, Tyson Limato

September 25, 2023

The Cancer AI project aims to develop an AI medical software engine to predict the progression of colorectal cancer in patients. We accomplish this goal with the following steps:

1. Condensing our input data to reduce computational cost and improve the efficiency of our AI tool. We do this through a method called embedding.

2. Mapping the evolution of COVID-19 and colorectal cancer to be provided as training data for our AI tool. We map the genetic development of both diseases through a biological concept known as a phylogeny tree.

3. We hypothesize that the evolution of COVID-19 and colorectal cancer can be used in our AI tool to improve performance. In technical verbiage, we are using a Long-Short-Term-Memory (LSTM) network to memorize state-action pairs of cancer cells of expert demonstrations to use for the IRL network.

4. We use a new AI technique called Inverse Reinforcement Learning (IRL) to predict the most dangerous evolutionary path for cancer mutations in order to ensure our AI tool can be used as a prognostic/diagnostic aid for real life medical patients to determine the cancer progression timeline. More explicitly, we utilize a reward function and a policy of IRL network to train the model to predict colorectal cancer progression.

This project requires a substantial amount of computer hardware and resources due to the advanced nature of the various AI techniques used. Therefore, this project uses Oracle Research Cloud resources to test the AI model. This project began in January 2023. As of September 25, 2023, we have analyzed the results of our embedding technique, yielding information on the quality of our condensed data as an input to the AI tool. As a prototype genome, we have created the phylogeny tree for COVID-19 and intend to create the phylogeny tree for colorectal cancer in the next stage of the project. If we constructed a phylogenetic tree of the human colorectal cancer genome on a local machine configuration such as an ordinary consumer laptop, it would take an estimated 7729 hours, or 88% of the year in computing. These computational demands alone makes the Cancer AI project infeasible without access to high performance computing. The LSTM model has been developed. We are currently devising our inverse reinforcement learning AI approach such that the internal function design satisfies the specific needs of the project.

In order to complete this research, each item mentioned above needs to be connected in a streamlined computational pipeline. The embedding technique needs to be thoroughly understood so the model accurately captures all relevant biological information and performs optimally in real-world applications. The inverse reinforcement learning implementation needs to be finalized. Finally, testing for the model needs to still be conducted on the Oracle Research Cloud. Therefore, prematurely ending the project at this stage yields no benefits for oncological study.

This project addresses the President of the University of Wyoming's objectives to act as a "catalyst for innovation" and to "provid[e] Wyoming and the world with agile and ethical computing [...] to address societal challenges that are inherently interdisciplinary."