Sprint Report: Team NSTC
#4: 31 Mar - 14 Apr

Backlog Progress

	Backlog Item	Progress/Remarks
Data Aggregation	Bash aggregation script	Continued debugging
Dataset Completion	HDF5 Implementation Run preprocessing	Testing complete Pending data aggregation
3D CNN	Debugging 9 2D->3D conversions GPU Memory testing	Expected completion Apr 16; Tensors are hard Pending completion of previous
2D CNN	Establish hardware for Inception Implement Inception V3 Figure out how to represent 3D image as 2D	Complete! Working on it. Tests being conducted.

Sprint Outcome Assessment

This sprint, the team again focused heavily on the individual assignments composing the many facets of this project. Building data aggregation and exploring additional methodologies to fine tune image preprocessing has been an underlying theme throughout the lifetime of the project, and continues to be a paramount part of the project as a whole. Bash has proved to be an effective, albeit unwieldy method of aggregating the data that will be the basis of the CNN/RNN training. Development in the implementation of data aggregation with Bash is ongoing.

We found new life in 2D transfer learning. After realizing the training and time constraints that come with the training of 3D models, we have decided to begin investigating a new approach in the form of 2D transfer learning. We will use Google Inception v3 and retrain the final layer to identify lung scans. If this succeeds, it will *drastically* reduce the time that we will need to spend training our neural network, at the cost of some accuracy. However this approach comes with a whole set of new challenges such as how are we going to represent a 3D structure as a 2D image without losings lots and lots of potential information. This is an issue we are still attempting to solve and may require a few tests to figure out what could yield the best results. We also had to prepare a new linux box to accomplish this task...

The completion of 3D code is behind schedule, but will return to a healthy pace with the completion of one senior thesis and a return to normalcy.

Project Assessment

We are on track to deliver two candidate models, one built around a 3D CNN and another built around a pretrained Inception CNN with a RNN wrapper. These will have variable, and likely insufficient, training time for production applications, but should have sufficient opportunity to demonstrate likely performance in such a setting. We are beginning work on the final deliverable, aiming for a general framework laid out in the first week of sprint 5.