Sprint Report: Team NSTC

#2: 24 Feb - 10 Mar

Backlog Progress

	Backlog Item	Progress	Remarks
Preprocessing	3D Function 2D Function Parameters: x,y,z,w; upscaling method; downscaling method (w=color_channels)		
Dataset	Expand the dataset ● ~~Flip axises~~ ● Blur Existing ● ~~Search for external sources~~	80%	Blurring created more problems than solutions at this point. See Outcome assessment for more details.
Computer Vision	Implement CNN function for: Inception, Alexnet, Highway Convolutional Parameters: input dimension; output dimension (maybe some internals?)		
Misc	RNN wrapper math (backprop over hybrid systems) Cloud compute setup Update requirements; test in virtual env Evaluate the use of custom filters		

Sprint Outcome Assessment

For this sprint the focus was to create the methodology for both 2D and 3D pre-processing. Along with such methodology, we came up with a process to implement a CNN for our project. The 2D pre-processing portion of the sprint had many road blocks, both in the development aspect and the overall usability. Due to this usability issue, additional research was needed for clarification. The research concluded that the 2D method may not be a viable way of approaching pre-processing. However, the 2D method for our project is still an alternative that is in consideration. The kaggle community has been an indispensible resource for the team's acquisition of pre-processing candidates. We have chosen to implement two of the top rated 3d

preprocessing methods as the forerunners of our initial implementation, and both will be in testing over the course of the coming sprints.

Dataset : Lack of data was a large problem that we needed to address. One option was to search online for more data published by medical sources. These can be found online on Kaggle, where other competitors are required to publish what they obtain. Another option, however, is to change the existing data that we already have, so that the neural network will recognize them as different images. Many challenges were introduced by this as well. For example, we discovered that in order to actually flip the data it must be standardized to begin with (Because matrix computations are confusing). This led to needing to coordinate both preprocessing code, and manipulation code. Another issue came about when attempting to blur the images. The methods used on each slice didn't produce a standard output, and when all put together created a distorted mess of data that likely won't be recognizable as lungs to the neural network. A new method of blurring will need to be tested.

Project Assessment

As our group mentioned in the previous sprint report, the size of the data was rather small and, as mentioned above, our team was successful in increasing the size of our data to allow our team to train our program more appropriately. Overall, progress is being made at a decent rate, despite this, however, based on our current sprint/backlog model, we are not quite up to the pace that we would like to be. This may be due to scope creep in which case a reorganization of our structure may need to be discussed in further detail in the future.

Complete Backlog

Planned Sprint	Backlog Items
Sprint 1 10 Feb - 24 Feb	1. Research a. ~~Familiarize with TF Learn, TensorFlow, TensorBoard, Serving~~ b. ~~Familiarize with dicom format, pydicom/alternate libraries, current interpretation methods~~ c. ~~Consider existing models~~ 2. Pre-Processing a. ~~Consider existing methods~~ **b. Implement/verify candidates**
Sprint 2 24 Feb - 10 Mar	1. C~~omputer Vision~~ a. ~~Consider existing~~ b. ~~Implement~~

	2. Time-Series ~~a. Consider existing~~ **b. Implement** 3. Probability a. Consider existing b. Implement
Sprint 3 20 Mar - 31 Mar	1. Connect Candidate Graphs 2. Test via TensorBoard/TPOT
Sprint 4 31 Mar - 14 Apr	1. Fixing Mistakes 2. Incorporating new developments
Sprint 5 14 Apr - 28 Apr	1. Deployment 2. Documentation