

Sprint Report: Team NSTC

#1: 10 Feb - 24 Feb

Backlog Progress

Backlog Item	Progress	Remarks
1a - TensorFlow familiarization	100%	
1b - Dicom familiarization	100%	
1c - Model familiarization	100%	
2a - Preprocessing model familiarization	75%	Still need to understand the use of filters to view bone vs. tissue in dicom slices.
2b - Preprocessing implementation	25%	Initial implementation will closely resemble Sentdex's layering of dicom slices. Kernel on Kaggle.
3a - Comp vision model research	99%	Continue to monitor arxiv.org and TF documentation for new models.
4a - Time series model research	99%	
5a - Probability model research	99%	

Sprint Outcome Assessment

This sprint focused primarily on attaining a team-wide baseline of familiarization with existing models and methods for a problem such as this. There was little original code written, but a strong list of existing methods was compiled. This will streamline implementation later in the process and allow the team to focus on experimentation. We are slightly behind an admittedly ambitious target, but have a realistic approach forward and can easily make up the time.

Project Assessment

A general consensus is emerging across the Kaggle community that the competition dataset may not be large enough to train a convolutional neural network as the central component of a classifier. Some additional data is available, although it is unclear if this will significantly improve results. To a large extent, we are waiting to see if the community generally tends towards another method or doubles down on the CNN. Aside from concern over data size, we are in a good position entering Sprint 2. Other than continuing to wait on community opinion, our

initial sprint wasn't completed due to 1c, 2a and 2b, found above. While attempting to complete these tasks, we found a great many candidates that we could potentially utilize in a model. Thus, we have added a list of potential candidates that we deem to have usability within our project. The list can be found at the end of this deliverable.

Complete Backlog

Planned Sprint	Backlog Items
<u>Sprint 1</u> 10 Feb - 24 Feb	<ol style="list-style-type: none"> 1. Research <ol style="list-style-type: none"> 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 <ol style="list-style-type: none"> a. Consider existing methods b. Implement/verify candidates
<u>Sprint 2</u> 24 Feb - 10 Mar	<ol style="list-style-type: none"> 3. Computer-Vision <ol style="list-style-type: none"> a. Consider existing b. Implement 4. Time-Series <ol style="list-style-type: none"> a. Consider existing b. Implement 5. Probability <ol style="list-style-type: none"> a. Consider existing b. Implement
<u>Sprint 3</u> 20 Mar - 31 Mar	<ol style="list-style-type: none"> 6. Connect Candidate Graphs 7. Test via TensorBoard/TPOT
<u>Sprint 4</u> 31 Mar - 14 Apr	<ol style="list-style-type: none"> 8. Fixing Mistakes 9. Incorporating new developments
<u>Sprint 5</u> 14 Apr - 28 Apr	<ol style="list-style-type: none"> 10. Deployment 11. Documentation

Candidate Models

PreProcessing	Dicom Analysis	Probability
<ul style="list-style-type: none"> ● 3D Projection ● Filters ● 2D Time Series ● Metadata 	Computer Vision (1 3D) <ul style="list-style-type: none"> ● Inception-v3/4 ● Alexnet ● Highway Convolutional 	<ul style="list-style-type: none"> ● None ● Linear Regression ● SVM ● Bayesian

<ul style="list-style-type: none">• External data	<p>Time Series (n 2D)</p> <ul style="list-style-type: none">• Simple RNN• LSTM• GRU• Bidirectional RNN	<ul style="list-style-type: none">• Clustering
---	---	--