PREDICTING COLON CANCER VIA DEEP LEARNING

JOSEPH WAUGH

Georgia Tech

CREATING THE NEXT®

PROJECT OUTLINE

- Project Goals and Requirements
- Team Roles & Contributions
- Research Based on Industry Problem
- Research Based on Gaps in Domain
- Demonstration of Application
- Project Status & Gantt Chart
- Discussion of Future Plans and Opportunities

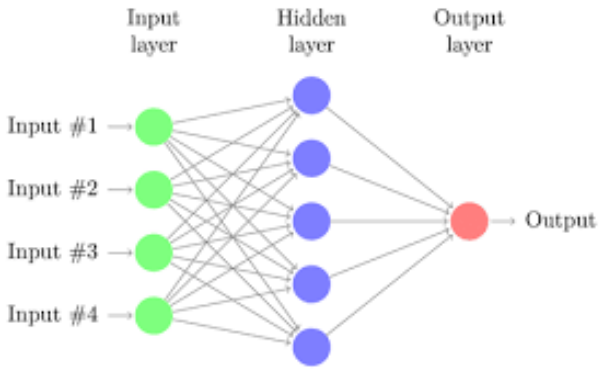
PROJECT GOALS & REQUIREMENTS

Goals:

- Create a neural network (NN) model which output a percentage likelihood whether the user has colorectal cancer
- Deploy the NN model via a Flask web application which allows a user to upload images of their CT scans, and receive feedback regarding their own scan

Requirements:

- Gain an understanding of the TensorFlow and Keras Python
- Secure a dataset to train our cancer-detecting model
- Deploy the web application via Heroku



TEAM ROLES & CONTRIBUTIONS

Project Number: Project #25 – Early Colon Cancer Detection in Men

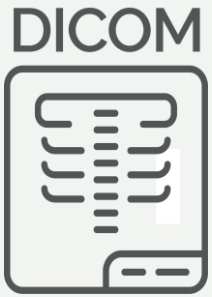
This project was listed as an **individual project** (and was completed individually), thus the following activities were completed individually:

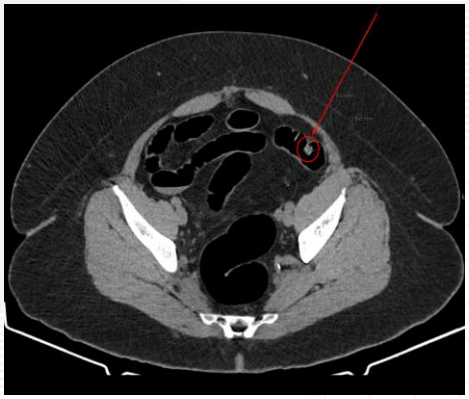
- Idea Generation
- Data Collection
- Model Development, Documentation & Deployment
- Web Application Creation

RESEARCH BASED ON INDUSTRY PROBLEM

The following topics were researched to better understand the colon cancer prevalence among men:

- Medical Imaging Standards (DICOM, CT Colonographies, Endoscopic Images)
- Availability of CT scan data (Kvasir Dataset*, The Cancer Imaging Archive)
- Opportunities with early cancer detection





**The Kvasir dataset was created using endoscopic imaging in the GI tract, where classes of "polyps" and "normal-cecum" were identified based on 500 cancerous and 500 non-cancerous images*

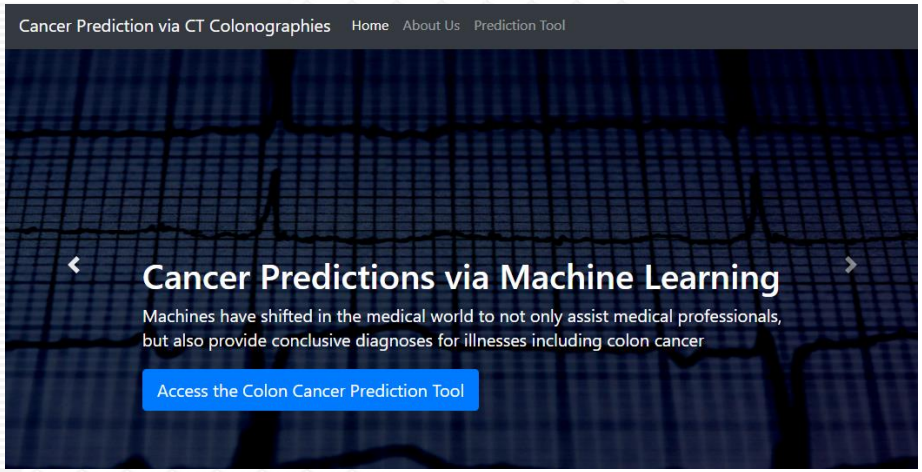
RESEARCH BASED ON GAPS IN DOMAIN

Given this topic is new based on the group's skillsets, the following topics required research:

- Neural Networks (i.e., TensorFlow & Keras Libraries)
- Heroku App Deployment (via Python Framework)

```
$ heroku git:clone -a polyp-predictor
$ cd polyp-predictor
```

```
$ git add .
$ git commit -am "make it better"
$ git push heroku master
```



RESEARCH

This project required additional research in the following areas to support the project development:

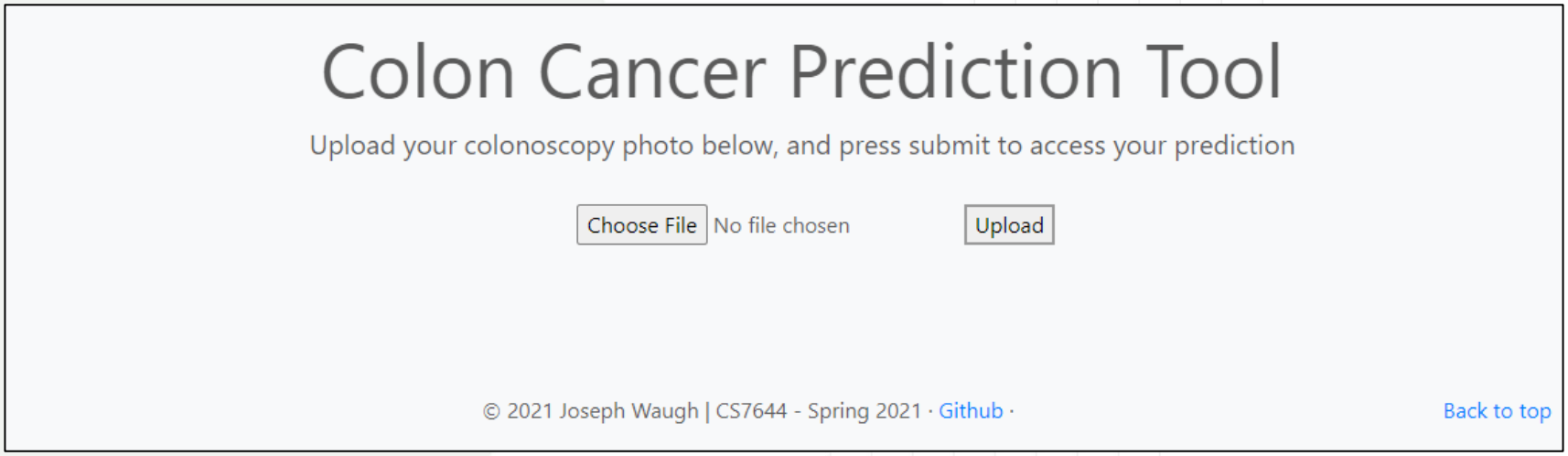
- Neural Networks (i.e., TensorFlow & Keras Libraries)
- DICOM Imaging in Deep Learning Models
- Heroku App Deployment (via Python Framework)

DEMONSTRATION OF APPLICATION

Georgia Tech

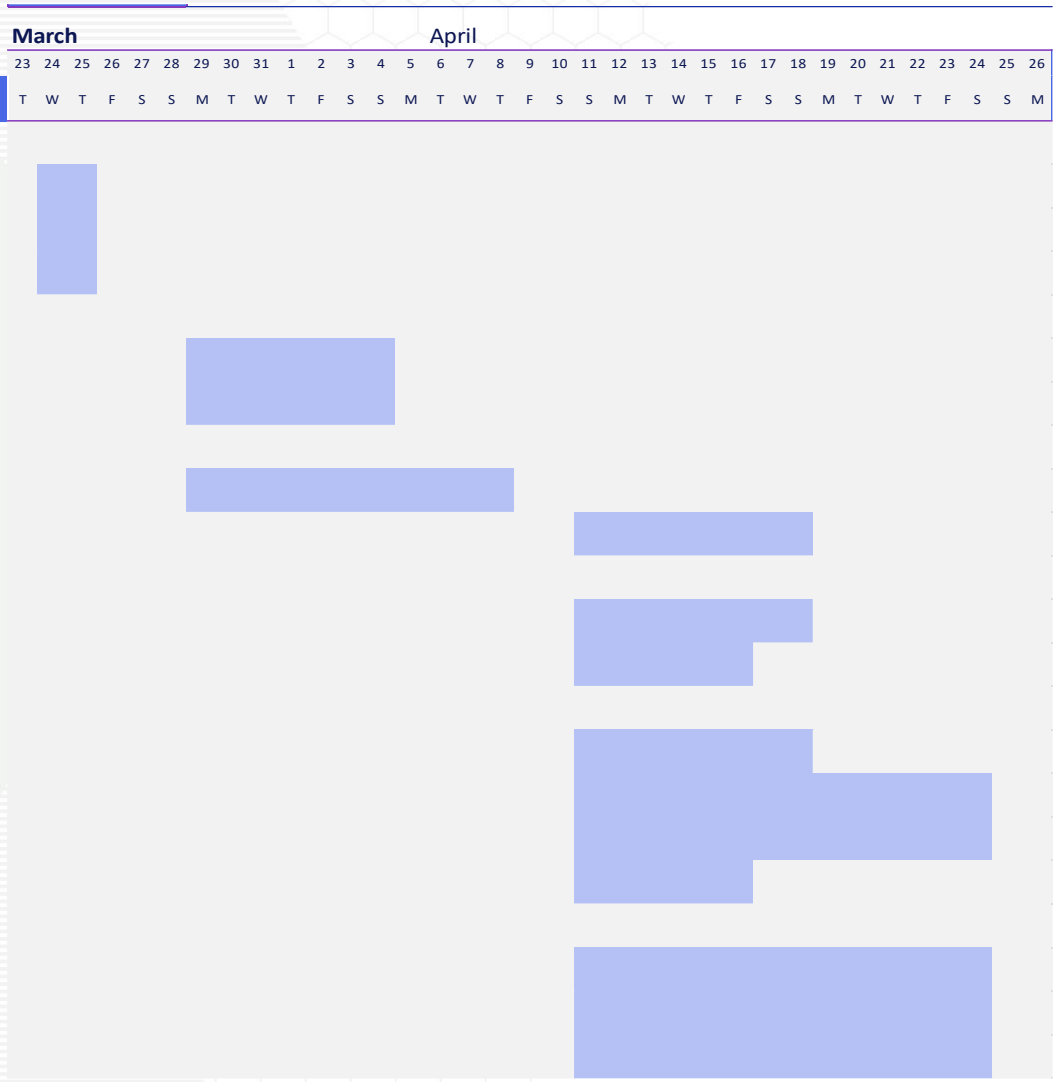
[Polyp Predictor Application Link]

Webpage Screenshot Page:



PROJECT STATUS/GANTT CHART

Milestone description	Start	End	Days
Week 11 (Week of 3/21)			
Download Cancer Imaging Archive Data via REST API	3/23/2021	3/27/2021	4
Create Flask Homepage for Application	3/23/2021	3/27/2021	4
Create Flask About Us Page for Application	3/23/2021	3/27/2021	4
Week 12 (Week of 3/28)			
Establish Framework for Keras Model	3/28/2021	4/6/2021	9
Process Cancer Imaging Dataset for Analysis	3/28/2021	4/6/2021	9
Week 13 (Week of 4/4)			
Process Cancer Imaging Dataset for Analysis	3/28/2021	4/10/2021	13
Train and Test Keras Model	4/10/2021	4/20/2021	10
Week 14 (Week of 4/11)			
Train and Test Keras Model	4/10/2021	4/20/2021	10
Create Dynamic Flask Diagnosis Page for Application	4/10/2021	4/18/2021	8
Week 15 (Week of 4/18)			
Train and Test Keras Model	4/10/2021	4/20/2021	10
Deploy Application via Heroku	4/10/2021	4/26/2021	16
Provide Documentation via Github	4/10/2021	4/26/2021	16
Create Dynamic Flask Diagnosis Page for Application	4/10/2021	4/18/2021	8
Week 16 (Week of 4/25)			
Deploy Application via Heroku	4/10/2021	4/26/2021	16
Provide Documentation via Github	4/10/2021	4/26/2021	16
Upload Video Regarding Usage	4/10/2021	4/26/2021	16



FUTURE PLANS AND OPPORTUNITIES

- To improve the model's scope, additional data can be used to train the CNN model
  - The model used the first of three versions of the Kvasir dataset*
- Additional classifiers for the polyp detection can be used to identify cancerous cells, normal cells, inflammatory diseases (i.e., esophagitis, ulcerative colitis, etc.)

**The Kvasir dataset contains three versions. For this research, version 1 was used (~1.2 GB), whereas the V2 (~2.3 GB) and V2.1 (~25.3 GB) are also available for public use*