Colon Disease Classification Using Machine Learning Techniques

Mitch Fulton
Dongyao Wang
Sushma Colanukudhuru
Chris Pillion

Project Motivation

Intestinal diseases are extensive and severe
- Affect between 60 and 70 million people worldwide
- Colorectal cancer is third most common kind of cancer, causing 50,000 deaths in 2016

Importance of early treatment
- Provides a 94% chance of survival
- Substantial portion of adults without regular colonoscopy checkups
  - Highlights the importance of accurate detection

Difficulty
- Detection is almost purely visual - computational methods not effective
- Uniformity of patterns & nonuniformity of the manifestation

Solution
- Leverage recent advancements in Machine Learning to visually detect diseases of the colon
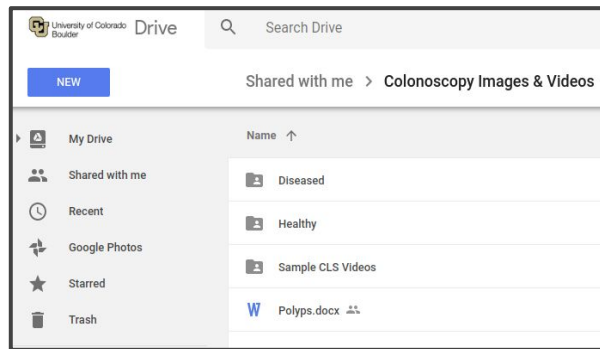
Model Considerations

Data Set
- Relatively small starting data set (~300 images)
- Labeled either healthy or by disease type
- Artificially increase the data set (scale, rotate, warp, etc)

Model Approaches
- Size of data set limits model choices
- Two main candidates:
  - Bayesian Learning and Support Vector Machines

Small data set
Highly dimensional inputs   →   Support Vector Machines

Given the limitations imposed by the size and type of our data set, our initial model will be constructed using Support Vector Machines (SVM)

Next Steps and Potential Obstacles

Access to datasets (AMTL)
- Not an ideal starting training set - results could vary immensely
- Solutions
    - Artificially increase dataset through image manipulation
    - Still single frame captures from videos
    - Solicit images from other medical research

Model Exploration
- Options for model choice limited
- Image recognition a dynamic and constantly evolving field
    - Want to look into recent advancements and techniques before settling on a final model