Using ML To Inspire A New Future For Pathology -

Colon Tissue Image Classification

Sprint 1

Yukie Kuang

Basic Workflow

Patient

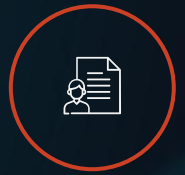
Has a lump

Doctor

Evaluates and orders a biopsy

Histologist

Processes, cuts sections, and stains biopsy with a preliminary stained slide



Pathologist

The Disease Detective

- orders and reads tests

It's benign!

What is Pathology, Histology, and Histopathology?

Pathologist

The Disease Detective

Pathology

Study of diseases

Histology

Study of the microscopic structure of tissues

Histopathology

Study of diseased tissue

Challenges in Histopathology for Pathologists

#1 **Subjectivity** Traditionally uses microscopes for slide

#2 **Scarcity** Physician shortages across the US

#3 **Decision Fatigue** Increased workload

Colorectal (Colon) Cancer on the Rise

- Also known as CRC
- Has been rising in young adults
- Rates have nearly doubled in younger adults age 50 or less (Dharwadkar, Zaki, & Murphy (2022))

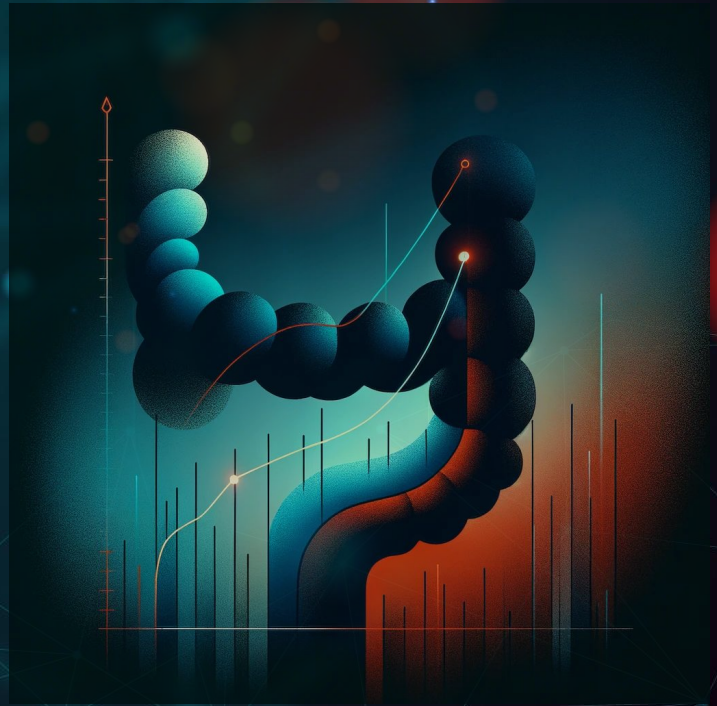


Image generated by OpenAI's DALL-E

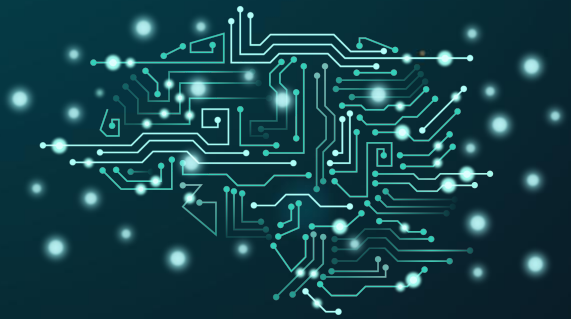
Image generated by OpenAI's DALL-E

The Impact of AI

- CRC is the most preventable, but least prevented cancer
- Early detection is the best prevention
- Increased volume of Colon biopsies
- **Role of AI and Deep Learning**:
  - Alleviate stress on pathology services
  - Offering researchers valuable insights
  - Feuling and aiding novel treatments for colon related diseases

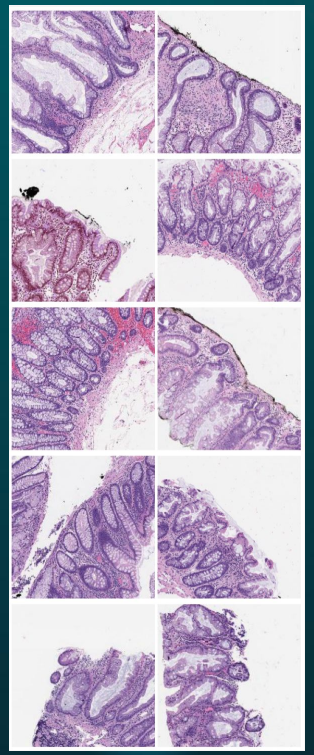
Project Focus

- Use Convolutional Neural Network ML to screen and classify colon tissue biopsy images
  - vital for early detection and treatment of colon-related diseases
  - understand the limitations of CNNs



Dataset

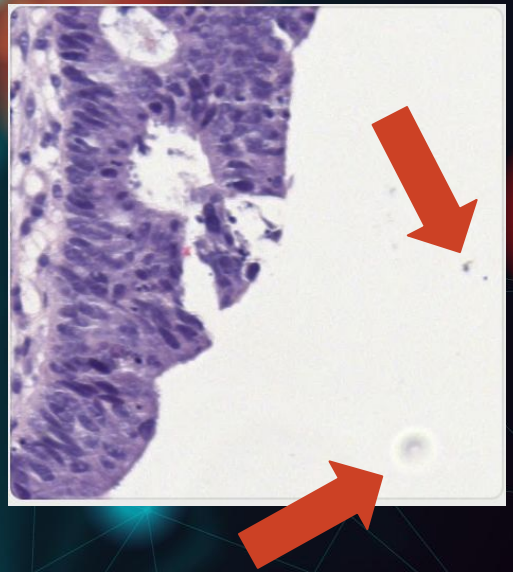
- Combination of dataset from a hospital in Chaoyang, Beijing on Github and a minimalist histopathology image analysis dataset (MHIST)
- 6,160 images and 3,152 images, respectively
- All colon tissue biopsies were stained with the same preliminary stain.



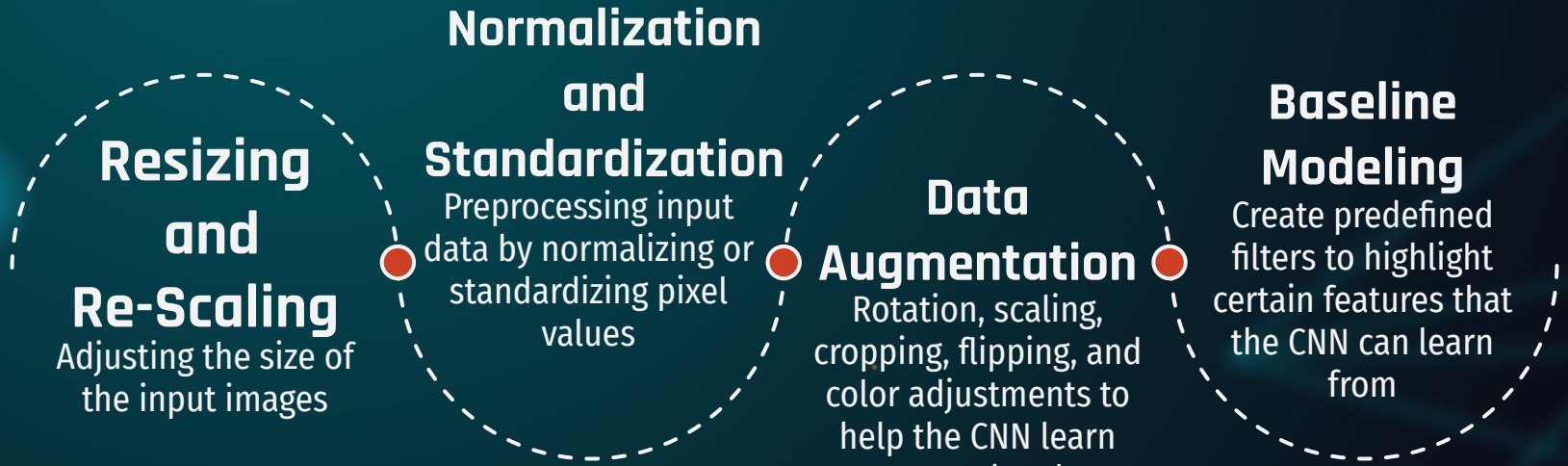
Preliminary EDA

Findings

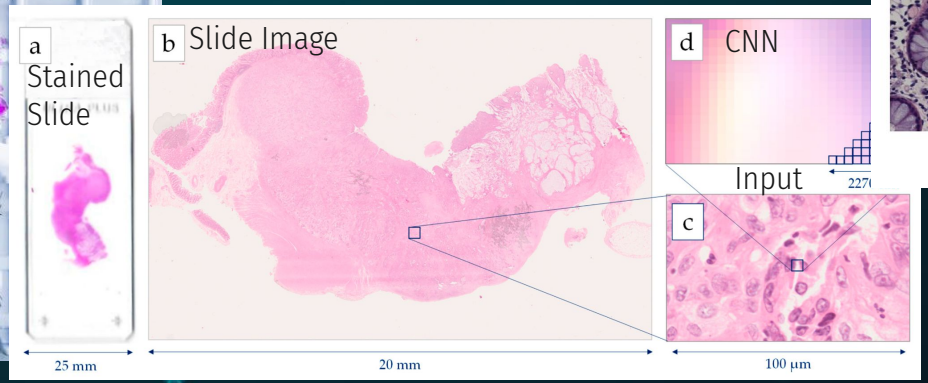
- Github dataset images have a uniform fixed size input size 512×512
- MHIST dataset images 224x224
- All images exist in the RGB Color space
- Images are representative of real-world noisy dataset scenario during digital image collection
  - Bubbles and artifacts on slides



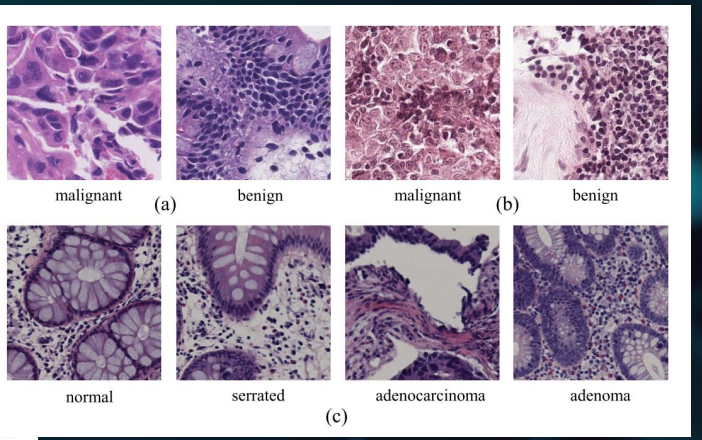
Next Steps


Resizing and Re-Scaling
Adjusting the size of the input images
Normalization and Standardization
Preprocessing input data by normalizing or standardizing pixel values
Data Augmentation
Rotation, scaling, cropping, flipping, and color adjustments to help the CNN learn more robustly
Baseline Modeling
Create predefined filters to highlight certain features that the CNN can learn from

Utilizing the physical to digital to create the automated



Davri. al. (2022)



Zhu et. al. (2022)

Citations:

Davri, A., Giannakeas, N., Birbas, E., Kanavos, T., Ntritsos, G., Tzallas, A., & Batistatou, A.(2022). Deep Learning on Images for Colorectal Cancer Diagnosis. In Encyclopedia. https://encyclopedia.pub/entry/21779

Dharwadkar P, Zaki TA, Murphy CC.(20022) Colorectal Cancer in Younger Adults. Hematol Oncol Clin North Am. 36(3):449-470. doi: 10.1016/j.hoc.2022.02.005.

Ruesch Center. (2021). Trends in colorectal cancer incidence among young adults. The Ruesch Center for the Cure of Gastrointestinal Cancers, Georgetown University. https://ruesch.georgetown.edu/youngadultcrc/

Wei, J. et al. (2021). A Petri Dish for Histopathology Image Analysis. In: Tucker, A., Henriques Abreu, P., Cardoso, J., Pereira Rodrigues, P., Riaño, D. (eds) Artificial Intelligence in Medicine. AIME 2021. Lecture Notes in Computer Science(), vol 12721. Springer, Cham. https://doi.org/10.1007/978-3-030-77211-6_2

Zhu, C., Chen, W., Peng, T., Wang, Y., & Jin, M. (2022). Hard Sample Aware Noise Robust Learning for Histopathology Image Classification. IEEE Transactions on Medical Imaging, 41(4), 881-894. https://doi.org/10.1109/TMI.2021.3125459