Early Detection of Colorectal Cancer

Supervised by:

Dr. Hanaa Bayomi
Dr. Ahmed Farouk
TA. Sara Ahmed

Implemented by:

IDs	Names
20188004	Esraa Sherif Mohamed
20188009	Anas Emad Abd-Elmoaty
20188022	Rawan Osama Bakr
20188036	Ali Mohamed Hussien
20188049	Mostafa Ahmed Mohamed

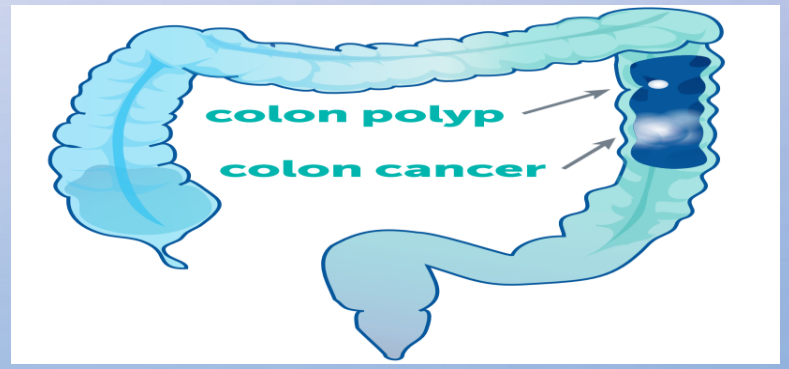
Agenda:

- Project Idea
- Problem Statement
- Related Work
- Methodology
- Specifications
- Time Plan
- Conclusion
- References

Project Idea:

- What is cancer?

- What is Colorectal Cancer?



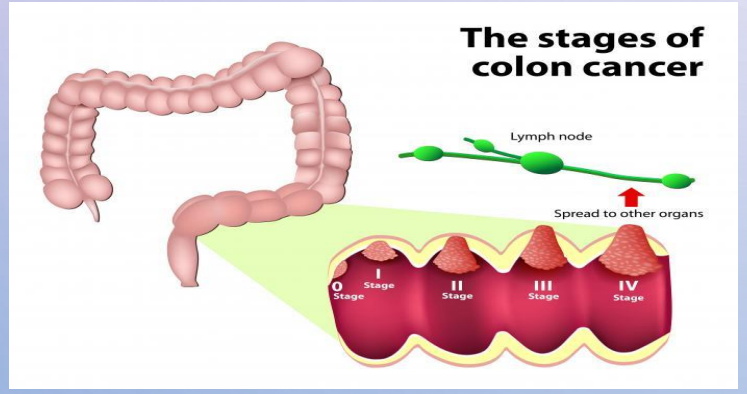
Project Idea:

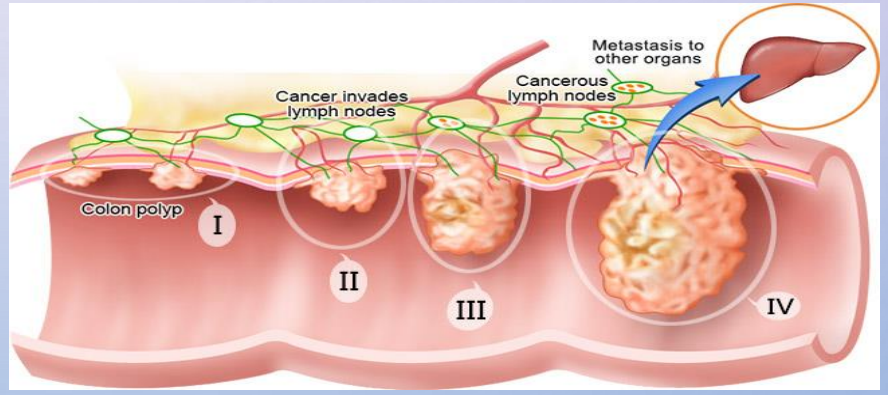
- Why we chose colorectal cancer?
  - The basic idea of the project came from the importance of early detecting colorectal cancer as Colorectal Cancer (CRC) is among the three most common cancers with more than 1.2 million new cases and about 600,000 deaths per year worldwide.
- What are the symptoms of Colorectal Cancer?
  - Blood in stool
  - Abdominal pain
  - Fatigue
  - Unexplained weight loss

Project Idea:

- What are the risk factors that increase the chance of colorectal cancer?

- What are the stages of Colon Cancer? (stages I, II, III, IV)





Problem Statement:

- Detection of Colorectal cancer is mostly based on traditional ways. These ways are chemical and experimental tests. If the signs and symptoms indicate colon cancer possibility, your doctor may recommend one or more tests.

- One of these tests is blood test but no blood test can tell you if you have colon cancer.

- Cancer is detected at late stage which decreases the probability of survival.

Problem Statement:

- Another test is using a scope to examine the inside of patient colon.

- Surgical tools are used to take tissue samples for analysis and remove polyps.

Problem Statement:

- Disadvantages:

  - There is a small risk of perforation the colon.
  - This test is not suitable for pregnant patients because of the use of radiation (another risk).

How to solve the problem?

- Problem Solution:

  - Doctors discovered that diagnosis of CRC have a relation with pathological finding includes **esophagitis, polyps and ulcerative colitis** .

  - The discovery of this pathological finding is achieved by **colonoscopy video**.

  - There is anatomical landmarks in GI such as **z-line , Pylorus and Cecum**.

How to solve the problem?

- Problem Solution:

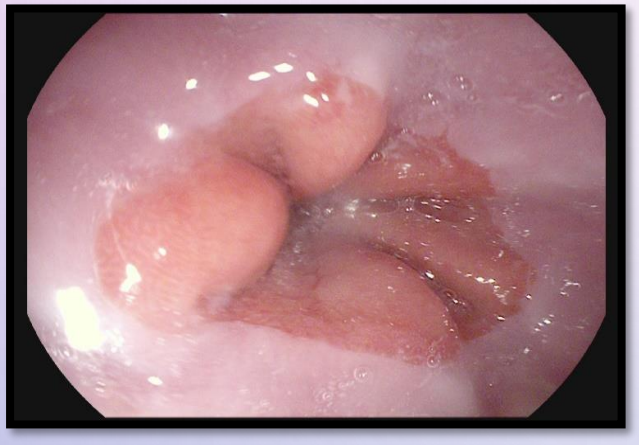
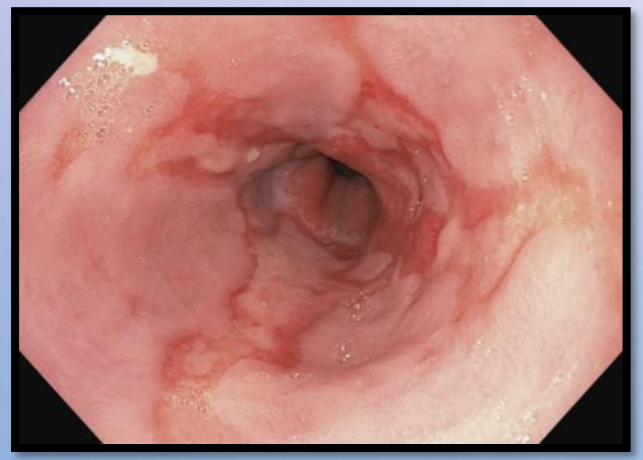
  - These anatomical landmarks refers to the location of pathological finding in GI.

  - The main objective of this project is early detection of colon cancer based on medical image processing on photos extracted from colonoscopy videos.

Dataset:

- We searched for dataset contain images from GI that covers anatomical landmarks that we have talked about it.
- We used kvasir, a dataset containing images from inside the GI that have 6000 images
- Each anatomical landmark has 2000 images that includes normal and abnormal case.
- We address a current issue in medical picture handling, the discovery of colorectal disease from colonoscopy videos.

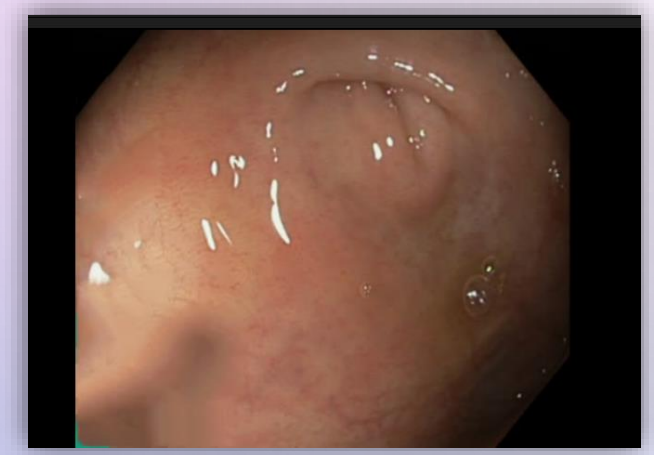
Dataset:

	***Z-line*** • The Z-line marks the transition site between the esophagus and the stomach • Containing 1000 images
	***Esophagitis*** • Esophagitis is an inflammation of the esophagus visible as a break in the esophageal mucosa in relation to the Z-line. • Containing 1000 images

Dataset:

	***Pylorus*** • The pylorus is defined as the area around the opening from the stomach into the first part of the small bowel (duodenum). • Containing 1000 images
	***Polyps*** • Polyps are lesions within the bowel detectable as mucosal outgrows. • Containing 1000 images

Dataset:

	***Cecum*** • The cecum the most proximal part of the large bowel. • Containing 1000 images.
	***Ulcerative Colitis*** • Ulcerative colitis is a chronic inflammatory disease affecting the large bowel. • Containing 1000 images.

Related Work:

- **Colorectal Cancer Detection Based on Deep Learning** **Aug 2020**
  - It was stated in this paper that a deep learning model (CNN) was used to detected CRC.
  - This deep learning approach produces median accuracy 89% .
  - The dataset consisted of 307 colorectal cancer-related digital slides from St. Paul's Hospital
  - 275 slides were randomly selected for training, the remaining slides were used for testing.

Related Work:

- **Deep Learning in Image Classification using ResNet Variants for Detection of Colorectal Cancer** **2020**
  - 3 models of deep learning(ResNet18,ResNet50,ResNet18&ResNet50) were used to detected CRC
  - The data set is 167 images (medical images that take by scanner)
  - ResNet 18 model produced accuracy 85% , when split data into 80% train & 20% test
  - ResNet 50 model produced accuracy 88% , when data split into 75% train,25% test
  - ResNet 18 & ResNet 50 produced accuracy 55% , when data split into 75% train ,25% test

Related Work:

➢ **Work Differences (Colorectal Cancer Detection Based on Deep Learning)**

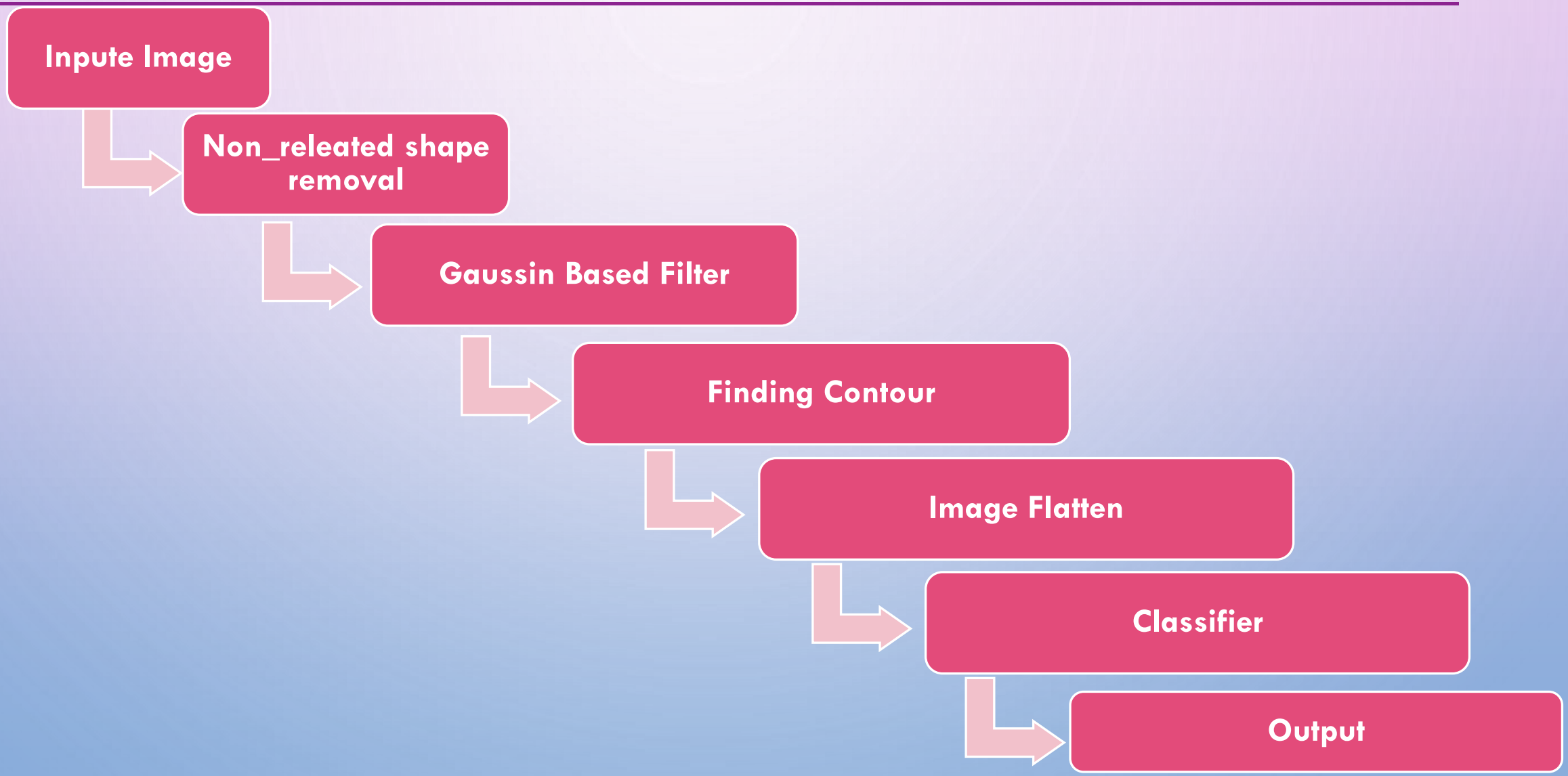
- Our data set is images extracted from colonoscopy video . Other projects have data set is images extracted slides digitized from clinical samples by Scanner
- Our project used CNN model and SVM model but also project1 model built with only CNN model , also project 2 built with 3 model of deep learning
- Our project has a large number of images up to 8000 unlike other projects

Methodology:

- Preferred programming language : python
- Reason of choice : as it is optimum when it comes to building machine and deep learning models.
- Used libraries :

Name	Usage
CV2	Reading And Processing Images
OS	To list The directory Content
Sklearn	Splitting the data , importing model and computing accuracy

Overview:


Inpute Image
Non_releated shape removal
Gaussin Based Filter
Finding Contour
Image Flatten
Classifier
Output

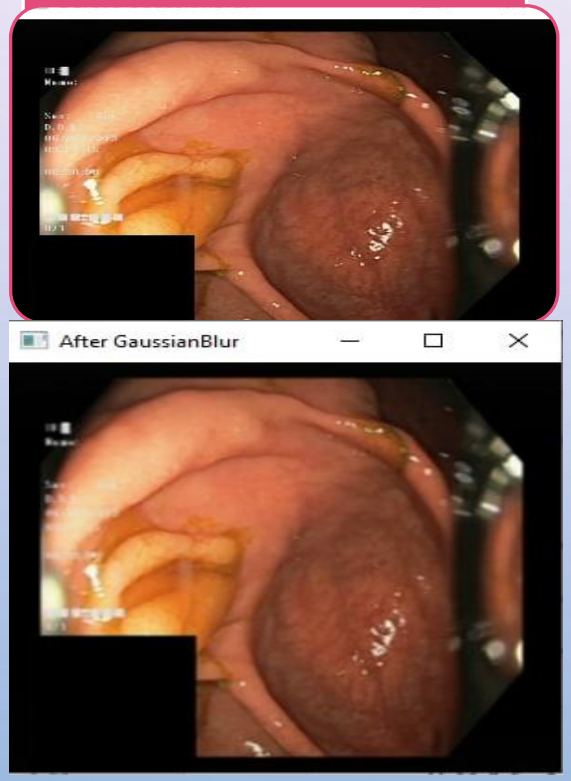
Methodology:

- Coding Steps .
- Firstly Reading images With its different anatomical landmarks each one have positive and negative class.
- Secondly Preprocessing images.

Methodology:

Removing the green square

Blurring images with GaussianBlur



Edge detection

Methodology:

Finding contours

Flatten images

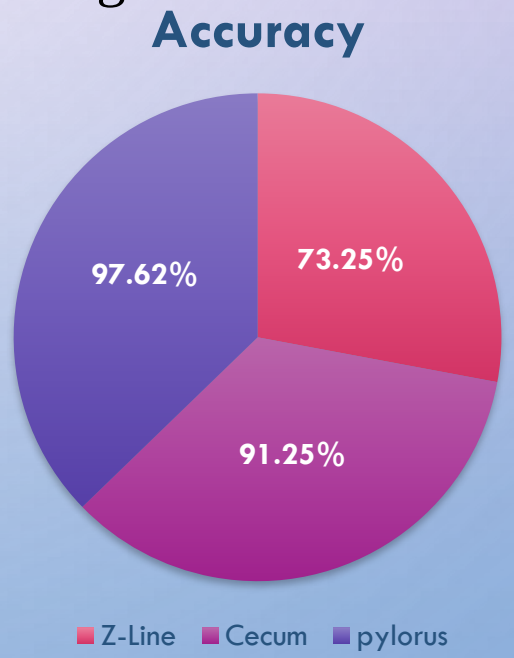
- Convert images into 1-d array

Classify image

- SVM
- ResNet50
- VGG19
- Model1
- Model 2

Evaluation:

- Split the data into train and test with Shuffle and it was used to build the model.
- One model was generated as a general model for all anatomical landmarks at once with accuracy 87% .
- Another model was built for each anatomical landmark resulting in three different model with a variety of accuracy.

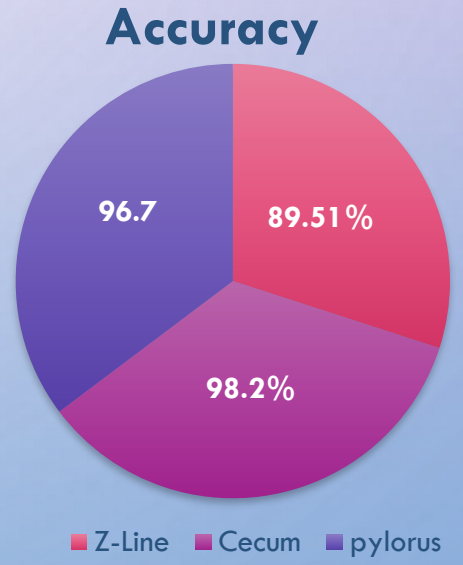


Evaluation:

- 4 CNN models were made
- Model had an accuracy of 73% with 30 epochs
- Model had an accuracy of 88% with 70 epochs
- Model with BUILT-IN VGG19 architecture had an accuracy of 87% with 25 epochs

Evaluation:

- Model with BUILT-IN RESNET50 architecture had an accuracy of 93% with 25 epochs
- Another model BUILT-IN RESNET50 was built for each anatomical landmark resulting in three different model with a variety of accuracy.



Time Plan:

Task Title	Status
Searching for project idea	Completed
Searching for data	Completed
Searching for reference project	Completed
Preprocessing of Data	Completed
Building of SVM model	Completed
Building of CNN model	Completed
Comparing between the accuracy of each model	Completed
Building Website	Completed

Conclusion:

- It is concluded that the common methods of CRC diagnosis has high risks.
- CRC can hardly be detected in it’s early stages.
- Therefore, proceeding with colonoscopy filming and using a model to classify the produced images to either CRC or healthy.
- However, the model classification cannot be considered as a 100% guaranteed result, hence, a specialized doctor must confirm if the result is valid or not.

Conclusion:

- It is expected that, the usage of the model would raise the survival rate of CRC patients and decrease the number of casualties as CRC has been detected in early stages.
- Given the aforementioned models (three specialized models VS on general model).
- It was foreseen that three specialized models is more accurate than a general model as each specialized model focuses only on pathological finding with its corresponding anatomical landmark using SVM model.

Conclusion:

- After tried fourCNN models that we mentioned before we noticed that our RESNET50 give higer accuracy
- Finally we put RESNET50 in our website

References:

- Xu, L., Walker, B., Liang, P.-I., Tong, Y., Xu, C., Su, Y. C., & Karsan, A. (2020, August 21). Colorectal cancer detection based on Deep Learning. Journal of pathology informatics.
- Kvasir. Simula Datasets - Kvasir. (n.d.). A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection
- Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021, February 19). Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Computer Science.