Cairo University

Faculty of Computers & artificial intelligence

Bioinformatics Program

جامعة القاهرة

كلية الحاسبات والذكاء الاصطناعي

المعلومات الحيويه

Early detection of colorectal cancer using machine learning & deep learning

---

This documentation submitted as required for the degree of bachelors in Computers and artificial intelligence.

Submitted by:

- *Mostafa Ahmed Mohamed* *Esraa Sherif Mohamed*
- *Ali Mohamed Hussein* *Rawan Osama Bakr*
- *Anas Emad Abd-Elmaoty*

Under Supervision of:

Dr. Hanaa Bayomi

Dr. Ahmed Farouk

TA. Sara Ahmed

July 2022

Cairo University

Faculty of Computers & artificial intelligence

Bioinformatics Program

جامعة القاهرة

كلية الحاسبات والذكاء الاصطناعي

المعلومات الحيويه

Early detection of colorectal cancer using machine learning & deep learning

July 2022

Acknowledgement

All praise and thanks to ALLAH, who provided us the ability to complete this work.

We would like to express our sincere gratitude and deep appreciation to our supervisors *Dr. Ahmed Farouk*,
Dr. Hanaa Bayomi , TA. Sara Ahmed who have supported us throughout our work with patience and knowledge.

Also, we would like to express our warmest appreciation to our families who are always providing help and support throughout the whole years of study. Without their support, we would have never had the chance to be who we are.

Abstract

Colorectal cancer (CRC) begins in the colon or the rectum. These diseases can likewise be called colon malignant growth or rectal cancer, contingent upon where they start. Colon malignant growth and rectal disease are regularly assembled in light of the fact that they share many highlights practically speaking. Colorectal disease is second reason for death in many communities additionally early conclusion have a more prominent possibility of survival. So colonoscopy represents a very important diagnostic modality for screening for colorectal cancer, we address a current issue in medical picture handling, the discovery of colorectal disease from colonoscopy videos. As per overall malignant growth measurements, colorectal disease is perhaps the most well-known disease. The most common way of screening and the expulsion of pre-malignant cells from the digestive organ is an essential undertaking to date. The traditional manual process is dependent on the expertise of the medical practitioner. We have Two phases in this project The first stage we utilized SVM model with an accuracy of 87 % in second stage we utilized deep learning CNN models with some accuracies.so, The main objective of this project is early detection of colon cancer based on medical image processing extracted from colonoscopy videos.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Word	Abbreviation
Colorectal cancer	CRC
Support Vector machine	SVM
Convolutional Neural Networks	CNN
Hypertext Markup Language	HTML
Cascading Style sheet	CSS
Java Script	JS
PHP Hypertext Preprocessor	PHP

Chapter One
Introduction

Introduction

- **What is Cancer?**

  The frame consists of about 15 trillion cells, and each day billions of cells wear out or are destroyed. Each time a cell is destroyed the body makes a brand new cell to exchange it, trying to create a cell that's an ideal copy of the cell that was destroyed because the replacement cell must be capable of performing the identical function because the destroyed cell.
  During the complex process of replacing cells, many errors occur.

  The chassis correct most mistakes, sometimes mistakes are made and not corrected, many of the uncorrected mistakes have little

effect on health.
If these mistakes allow the newly made cell to divide independent of the checks and balance that control normal cell growth, that cell can begin to multiply in an uncontrolled manner, when this happens, a tumor can develop.

- **What is colorectal cancer?**

Colorectal cancer is the cancer of colon and rectum. It starts when the process of the normal replacement colon lining cells deviate. Health care professionals are certain that colorectal cancer is not contagious (a person cannot catch the disease from a cancer patient). Some people are more likely to develop colorectal cancer than others.

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

1. Age :

Increasing age is the main risk factor for colorectal cancer. Around 90% of colorectal cancers are diagnosed after age 50.

2. Race:

African Americans have a higher incidence of colorectal cancer than people of other races.

3. Diet and colorectal cancer:

Diets high in fat have been shown in numerous research studies to predispose people to colorectal cancer.

In countries with high colorectal cancer rates, the fat intake by the population is much higher than in countries with low cancer rates. It is believed that the digestion of fat that occurs in the small intestine and the colon leads to the formation of cancer-causing chemicals (carcinogens). Likewise, research studies also reveal that diets high in vegetables and high-fiber foods such as whole-grain breads and cereals contain less fat that produces these carcinogens and may counter the effects of the carcinogens. Both effects would help reduce the risk of cancer.

4. Inherited syndromes:

Genetic syndromes passed through generations of your family can increase your risk of colon cancer.

5. A sedentary lifestyle:

If you're inactive, you're more likely to develop colon cancer. Getting regular physical activity may reduce your risk of colon cancer.

6. Diabetes:

People with diabetes and insulin resistance have an increased risk of colon cancer.

7. Obesity:

People who are obese have an increased risk of colon cancer and an increased risk of dying of colon cancer when compared with people considered normal weight.

8. Smoking:

People who smoke may have an increased risk of colon cancer.

9. Alcohol:

Heavy use of alcohol increases your risk of colon cancer.

10. Radiation therapy for cancer:

Radiation therapy directed at the abdomen to treat previous cancers increases the risk of colon and rectal cancer.

- **<u>What are the symptoms of Colorectal Cancer?</u>**

Symptoms of colorectal cancer are numerous and nonspecific. **They include:**

- Fatigue,
- Weakness,
- Shortness of breath,
- Change in bowel habits,
- Narrow stools,
- Diarrhea or constipation,
- Red or dark blood in stool,
- Weight loss,
- Abdominal pain,
- Cramps or bloating.

- **What are the stages of Colon Cancer?**

  - **Stage I:**

    The cancer has grown through the superficial lining (mucosa) of the colon or rectum but hasn't spread beyond the colon wall or rectum.

  - **Stage II:**

    The cancer has grown into or through the wall of the colon or rectum but hasn't spread to nearby lymph nodes.

  - **Stage III:**

    The cancer has invaded nearby lymph nodes but isn't affecting other parts of your body yet.

  - **Stage IV:**

    The cancer has spread to distant sites, such as other organs for instance, to your liver or lung.

- **What is Machine Learning?**

  is a field of understanding and building methods that 'learn', that methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence.

  Machine Learning Algorithm:

  algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being

explicitly programmed to do so.
It used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks

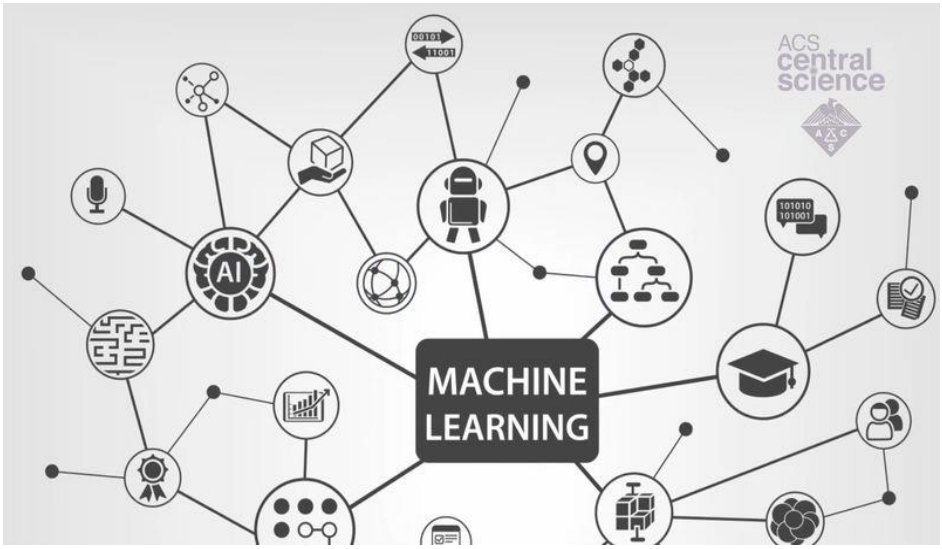
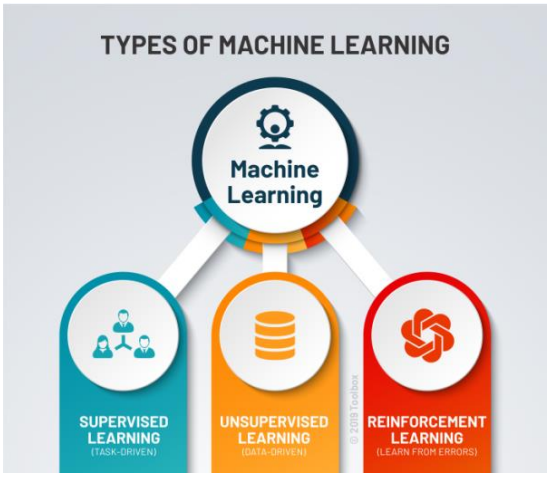


Fig1.1 Machine learning

- **Machine Learning Structure:**



Within the field of machine learning, there are two main types of tasks: supervised, and unsupervised.

The main difference between the two types is that supervised learning is done using a ground truth**,** or in other words, we have

Fig 1.2 types of ML

Prior knowledge of what the output values for our samples should be. Therefore, the goal of supervised learning is to learn a function that, given a sample of data and desired outputs, best approximates

the relationship between input and output observable in the data.

Unsupervised learning, on the other hand, does not have labeled outputs, so its goal is to infer the natural structure present within a set of data point

Machine Learning can help in:

Prediction of cancer susceptibility – risk assessment prior to occurrence.

Prediction of cancer recurrence – likelihood of redeveloping.

Prediction of cancer survivability – life expectancy,

survival, progression, tumor-drug sensitivity.

Supervised Learning:

we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.
Supervised learning problems are categorized into "regression" and "classification" problems.
In a regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function.
In a classification problem, we are instead trying to predict results in a discrete output.
In other words, we are trying to map input variables into discrete categories

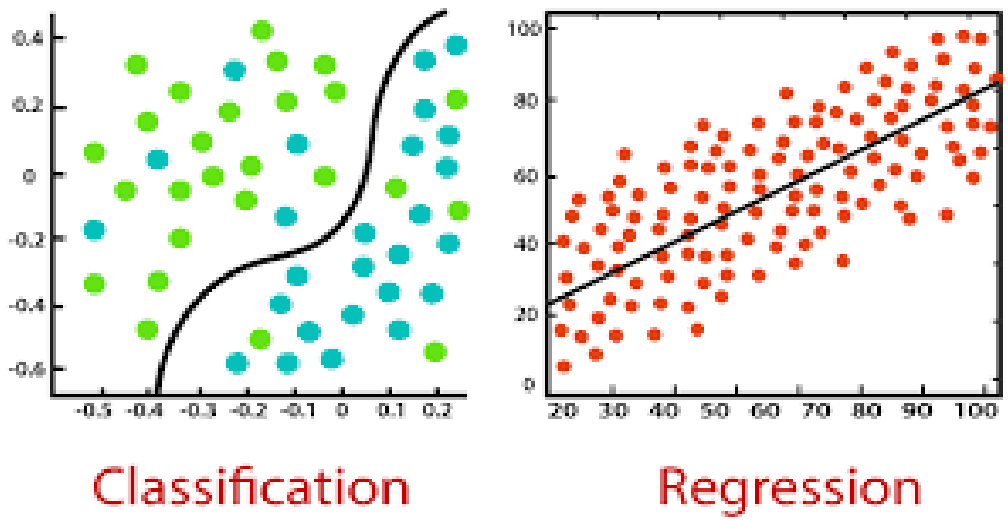


Fig1.3 Supervised Learning

Unsupervised Learning:

Unsupervised learning, on the other hand, allows us to approach problems with little or no idea what our results should look like.
We can derive structure from data where we don't necessarily know the effect of the variables. θWe can derive this structure by clustering the data based on relationships among the variables in the data.
With unsupervised learning there is no feedback based on the prediction results, i.e., there is no teacher to correct you

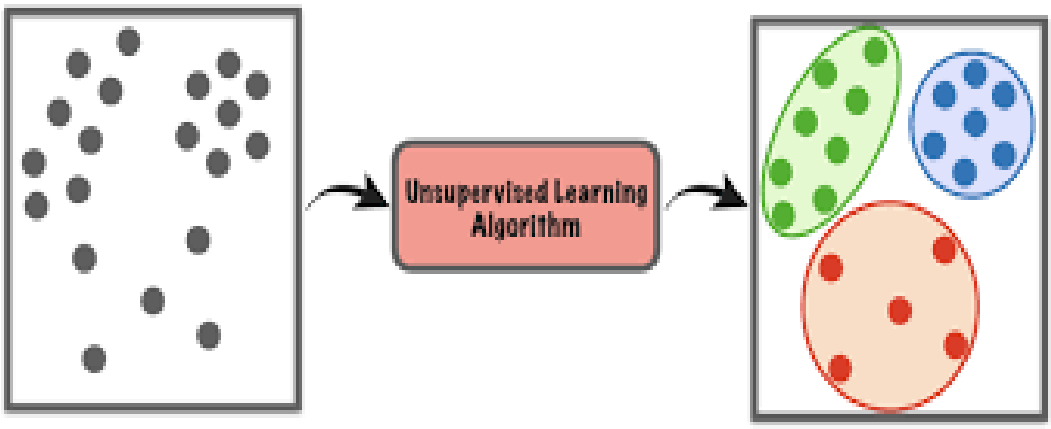


Fig 1.4

1.1 Motivation

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. The diagnosis of colorectal cancer at late stage makes the treatment more difficult and decreases to less than 10% chance of survival, on the contrary, early diagnosis of CRC is easier to treat and patient has 90% chance of survival. Using the medical pictures of various areas in the colon

Machine learning and deep learning models that predict whether the given picture is positive(has cancer) or negative(has no cancer)

1.2 Problem Definition

Currently, all detection of colorectal cancer based on traditional ways which are chemical and experimental studies as if your signs and symptoms indicate that you could have colon cancer, your doctor may recommend one or more tests and procedures, including:

Using a scope to examine the inside of your colon:

Colonoscopy uses a long, flexible and slender tube attached to a video camera and monitor to view your entire colon and rectum. If any suspicious areas are found, your doctor can pass surgical tools through the tube to take tissue samples for analysis and remove polyps.

Blood tests:

No blood test can tell you if you have colon cancer. But your doctor may test your blood for clues about your overall health, such as kidney and liver function tests. So, they detect cancer at late stage which decreases the probability of survival

1.3 Project Objective

The main objective of this project is early detection of colon cancer based on colonoscopy

Using a machine learning approach (SVM) to implement a classifier that can be trained on the labeled data, then it can classify the images into three different colonic locations and predict the case of that area whether it has a cancer or not

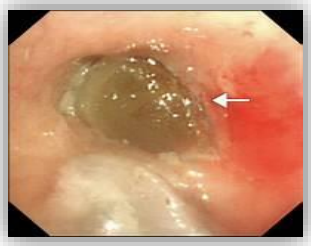
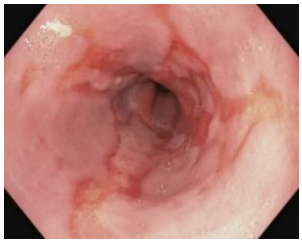
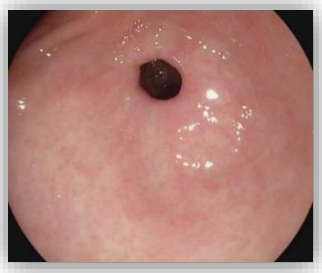
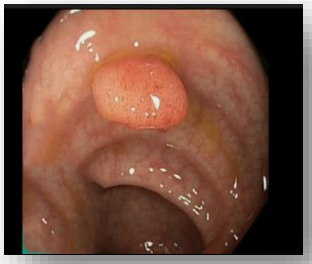
Using a deep learning approach (CNN) to implement model that can be trained on labeled data , so it can classify the images into three different colonic locations and predict the case of that area whether it has a cancer or not

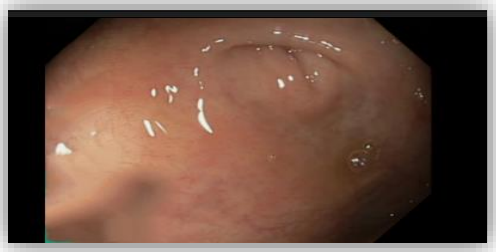
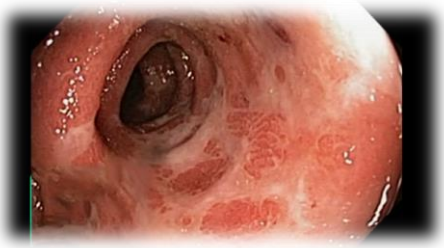
1.4 Gantt chart of project time plan

Fig1.5 Gantt Chart

1.5 Project development methodology

Our project is about early detection of colorectal cancer first by machine learning specifically SVM model, second by deep learning and we used some models with different layers to give different accuracy. we present Kvasir, a dataset containing images from inside the digest, Kvasir is important for research on both single- and multi-disease computer aided detection. By providing it, we invite and enable multimedia researcher into the medical domain of detection and retrieval. The Kvasir dataset consists of images, annotated and verified by medical doctors (experienced endoscopists), including several classes showing anatomical landmarks, pathological findings or endoscopic procedures in the digest, hundreds of images for each class such as:

Z-line	***Esophagitis***	***Pylorus***	***Polyps***
The Z-line marks the transition site between the esophagus and the stomach. 	Esophagitis is an inflammation of the esophagus visible as a break in the esophageal mucosa in relation to the Z-line. 	The pylorus is defined as the area around the opening from the stomach into the first part of the small bowel (duodenum). 	Polyps are lesions within the bowel detectable as mucosal outgrow 

Cecum	***Ulcerative Colitis***
The cecum the most proximal part of the large bowel 	Ulcerative colitis is a chronic Inflammatory disease affecting The large bowel. 

First phase:

we used is a machine learning approach called support vector machine (SVM) that is Supervised learning is typically done in the context of classification, when we want to map input to output labels, or regression, when we want to map input to a continuous output, so we prepared the dataset before classify

Firstly by reading images with its different anatomical landmarks each one have positive and negative class, and it's for linearly separable binary set ,main Goal to design a hyper plane that classify all training vectors into two classes ,The best model that leaves the maximum margin from both classes. the two classes labels +1 (positive examples and -1 (negative examples).

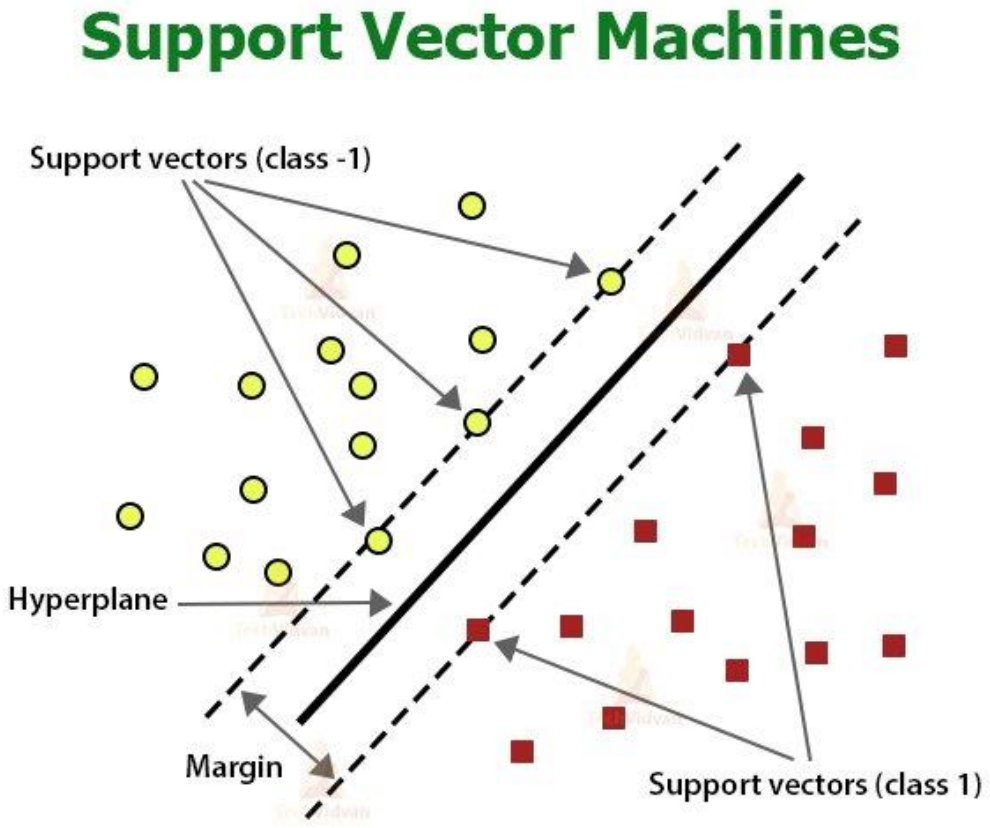


Fig 1.6

Secondly Preprocessing images

Removing the green square

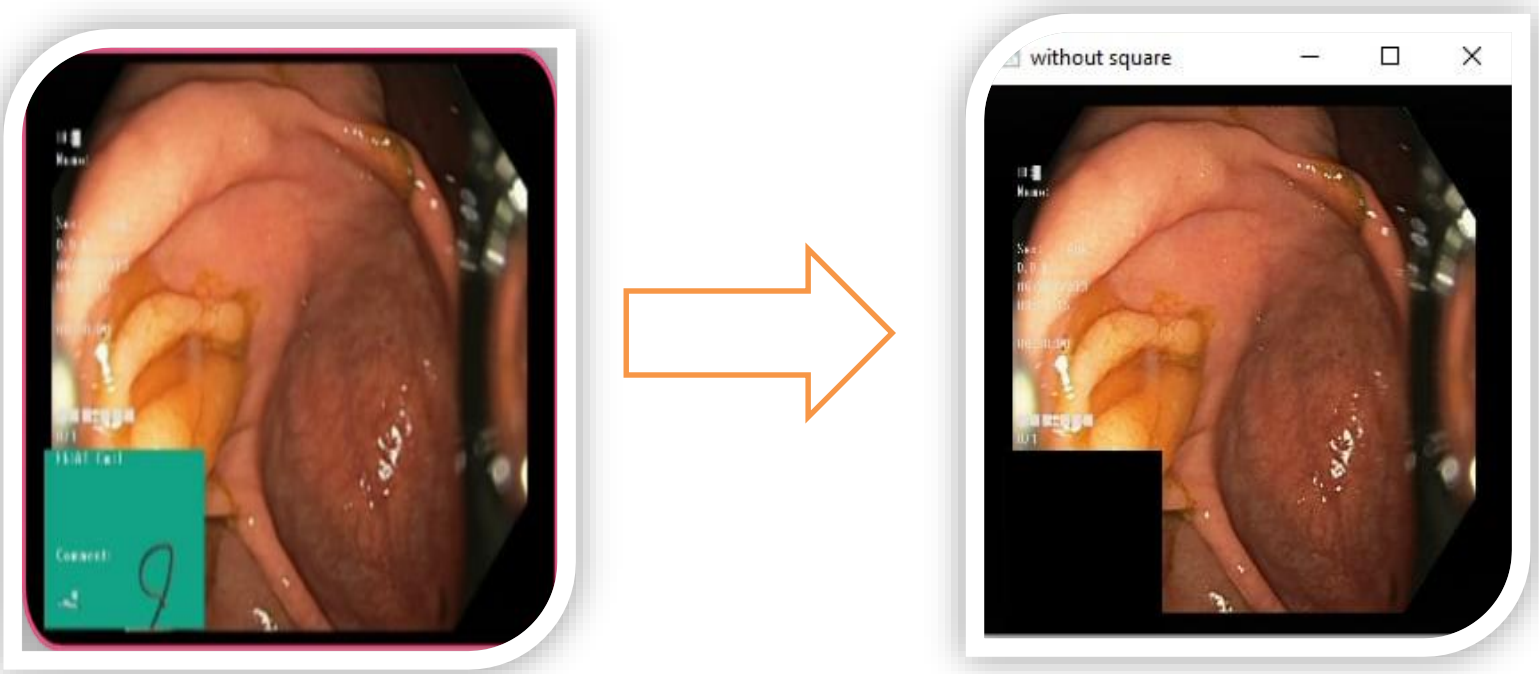
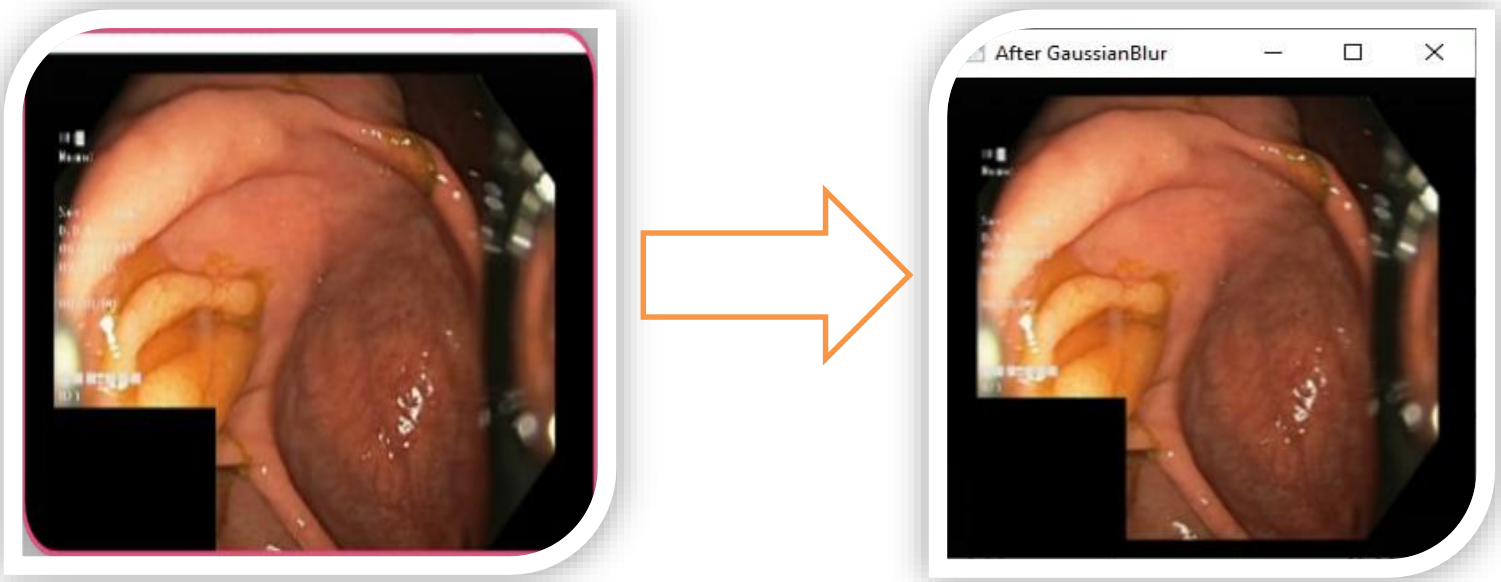


Fig 1.7

Blurring images with Gaussian Blur



Edge detection

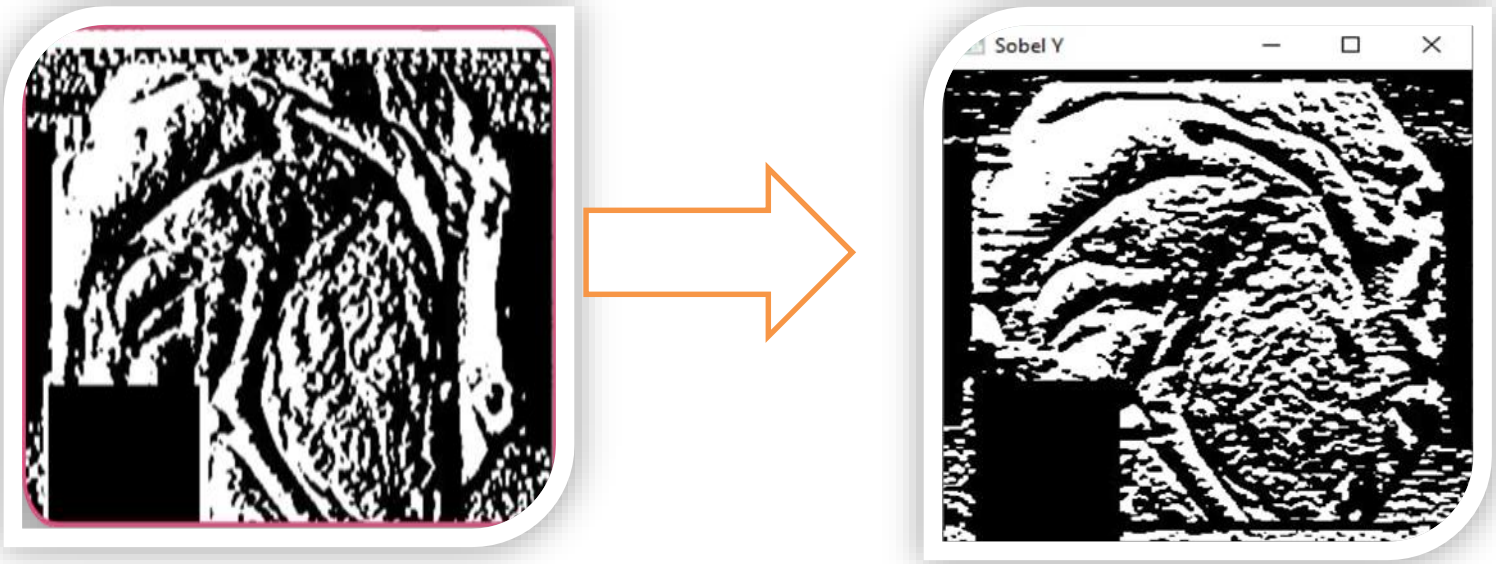


Fig 1.8

Finding contours

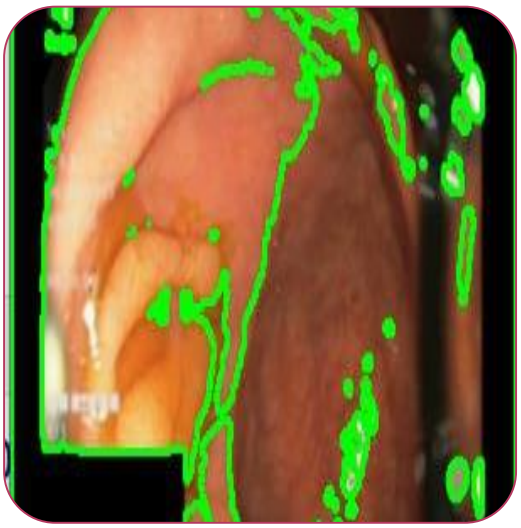


fig 1.9

Flatten images
Convert images into 1-d array

then the images begin to classify using SVM model and result in accuracy 87%

Second stage :

we used deep learning model called Convolutional Neural Networks(CNN) Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain CNN, is a deep learning neural network designed for processing structured arrays of data such as images. Convolutional neural networks are widely used in computer vision and have become the state of the art for many visual applications such as image classification, and have also found success in natural language processing for text classification, so we perform preprocessing on data like we mentioned above and used different structures of CNN models. CNN is a feed-forward network that can extract topological properties from an image.
Like almost every other neural networks they are trained with a version of the back-propagation algorithm. CNN are designed to recognize

visual patterns directly from pixel images with minimal preprocessing. They can recognize patterns with extreme variability (such as handwritten characters).

CNN is a deep learning neural network designed for processing structured arrays of data such as images. CNN are widely used in computer vision and have become the state of the art for many visual applications such as image classification, and have also found success in natural language processing for text classification.

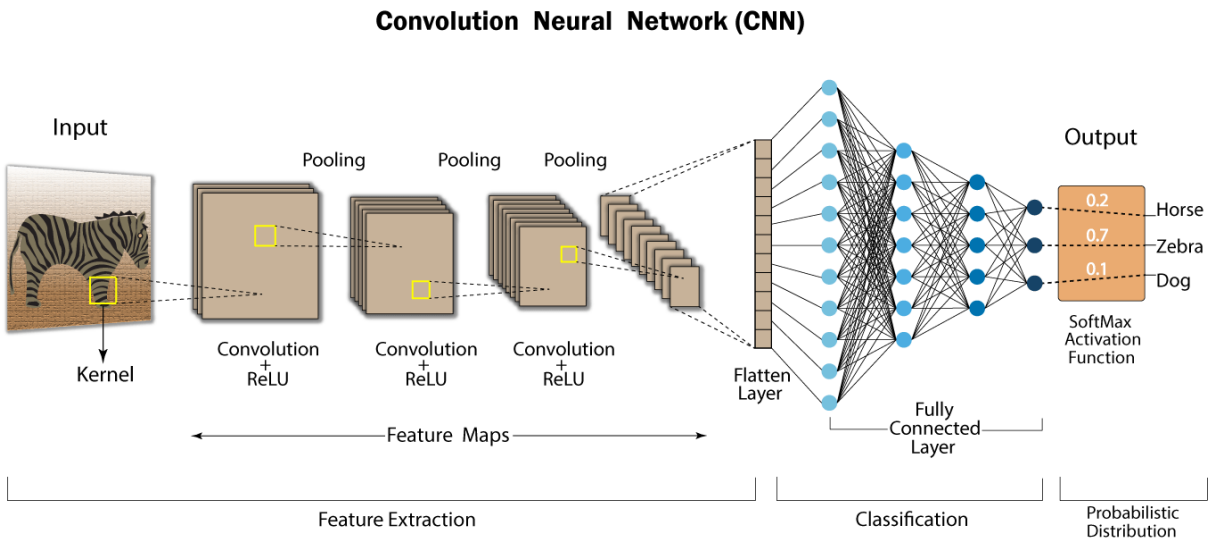


Fig 1.10

First we build a structure of CNN model from scratch, that contains conv2D with filter size 32 , max Pool 2*2, conv2D with filter size 64 and max Pool 2*2 , conv2D with filter size 128 and max Pool 2*2 ,

conv2D with filter size 256 and max Pool 2*2 , conv2D with filter size 512 and max Pool 2*2 then flatten and Dense, and we trained this model with the data that contains 6,000 images splitting it into 60% train data ,40% test data with 30 epochs.This model result in accuracy 73%

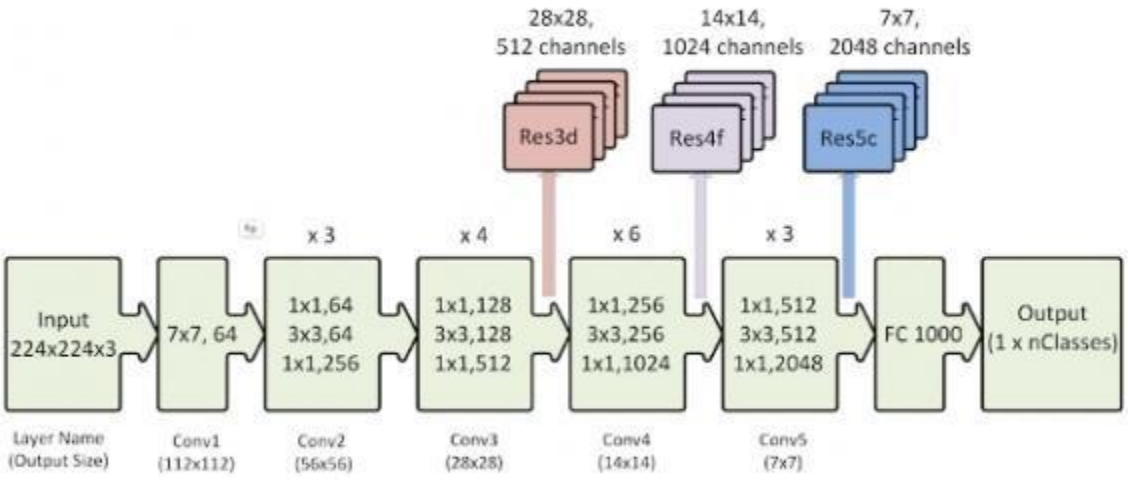


Fig1.11

For improving this accuracy we tried building another complex model with more layers and more number of filters: we build conv2D with filter size 200 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 240 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 280 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 280 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 320 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 360 ,maxPool 2*2 and Dropout with rate equals 0.2, conv2D with filter size 400 ,maxPool 2*2 and Dropout with rate equals 0.2 ,Flatten ,three layers of Dense with size 512,Dropout with rate equals 0.2 and Dense, to fit with data splitting the data also into 60% train ,20%validate and 20%test but with 70 epochs

and this model give accuracy 88%

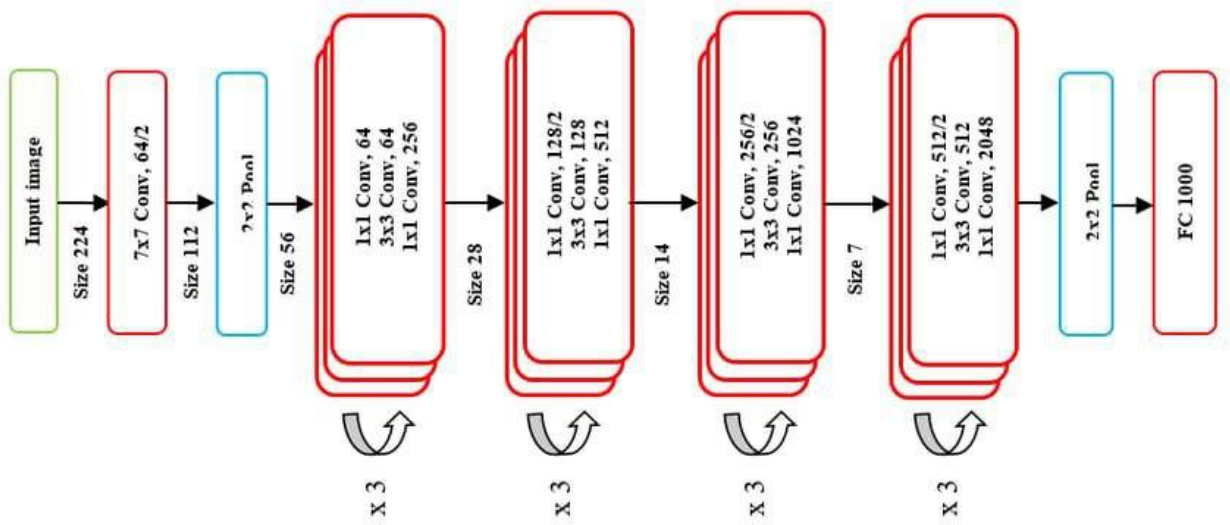


Fig1.12

We tried using built-in models in the hope of getting higher accuracy
First we used the built-in model VGG19 with 25 epochs and this model let to accuracy 87%

Finally:

we used built-in ResNet50 hoping of getting a high accuracy with 25 epochs and this model let to accuracy 93%

1.6 The used tools in the project

The Software tools that used in project:

Operating system	Applications
• Mac • Windows	• Visual studio Code • HTML • PHP • CSS • JS • PyCharm • Python 3.10 • UsbServer • MAMPServer • PHPMyAdmin • Skitlearn • Opencv • Tensorflow • Keras

1.7 Report Organization

Chapter Two: Related Work

In chapter two, we will establish other work associated with our research. There, we show that different ways to achieve our goals. We will declare the authors of these methods, the classification method they use in their project, and off course the accuracy.

Chapter Three: System Analysis

In chapter three, we will talk about project specification. It will contain descriptions of how the program will be used from a user perspective and performance details such as usability, reliability, and stability. In addition, illustrate a use case diagram that emphases our program.

Chapter Four: System Design

The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces using some diagrams. System Component Diagram, System Class Diagrams, Sequence Diagrams, Project ERD, System GUI Design

Chapter Five: Implementation and Testing

This phase includes screen shots of training and testing of (SVM,) model, CNN model
Screenshots of the system running and samples of the applied test cases.

Chapter Two
Related work

Colorectal Cancer Detection Based on Deep Learning :

The traditional idea:

was for a specialized pathologist to make an assessment
of(H&E)-stained tissue sections by microscopy
of multiple samples taken from one patient
This is step need more attention to detail .
In addition diagnoses depended on a individual pathologist's experience

Solution: using CNN model and using digitized H&E-stained histology slides
Accuracy is 89%
Dataset:

307 (85 normal & 222 colorectal cancer tissue)
275 slides (76 normal and 199 containing colorectal cancer) were randomly selected for training; the remaining slides were used for testing (9 normal and 23 colorectal cancer)

Work limitations:

-the preparation needed to build such model, from samples taken from patients' colon (health risk) to digitizing the slides is intensive and time consuming -Data amount is very small

Deep Learning in Image Classification using ResNet Variants for Detection of Colorectal Cancer

The traditional idea:

A pathologist collects biopsies
From all parts of colon, cells to identify cancer
They examine the tissue so that they can tell what stage this cancer is in.
This is done by examining the morphology of the cells and observing the changes that occur in them, as well as in the components of tissues visually under a microscope**.**
However, pathologists often disagree with regard to classifications of cancer, it is not enough to rely on the expertise of scientists alone.

Solution: Medical imaging presently serves to distinguish the diseases in various components of the body. And 3 models of deep learning (ResNet18,ResNet50,ResNet18&ResNet50)

Accuracy is 85%

Dataset:

-The data consist of 165 images with 74 benign tumors images and 91 malignant tumors images.
-This data was attained with Scanner

Work limitations

ResNet 50 take more time than ResNet 18 to deal with large data
Data amount is very small to depend on to prepare model

Work Differences :

- our project used CNN model and SVM model, but project 1 used only CNN model , project 2 used 3 models ResNet18,ResNet50,(ResNet18&ResNet50)

- in our project have 8000 images on another hand project 1 have 307 images, project 2 have 165 images [very small data]

- our data set used colonoscopy video , other projects used slides digitized from clinical samples by Scanner

Chapter Three
System Analysis

3.1 Project specification

3.1.1. Functional requirement

- Registration: If this is the first time for the user to use the website, he/she will register.
- Log in: If the user already registers before, the user will log in to the app with his user's name and password
- profile: the user can show his history, and he could back to the history and check the results and accuracy.
- Adding information for new Patient Or choose existing one
- Upload image as input
- image processing
- Determine if this image have cancer or not by model
- Output the classification and accuracy.

3.1.2. Non-functional requirement

- Usability
  The site is easy for use
- Security
  Login Authentication
  The site is secure, because no one can access without e-mail

- Reusability
  The system can be used in another medical purpose without a lot of change.

- Response Time
  The site doesn’t take more than 5 second to classify the images

- Robustness
  The system can deal with errors from the user during execution without crashing.
  The deployed model doesn’t suffer from overfitting or underfitting. The difference between train and test accuracy less than 1%.

- Availability
  The site is available 24 hours

- Maintainability
  It’s easy to deploy new model better than the model used in app.
  The system can be easily repaired and maintained

3.2. Use case Diagrams

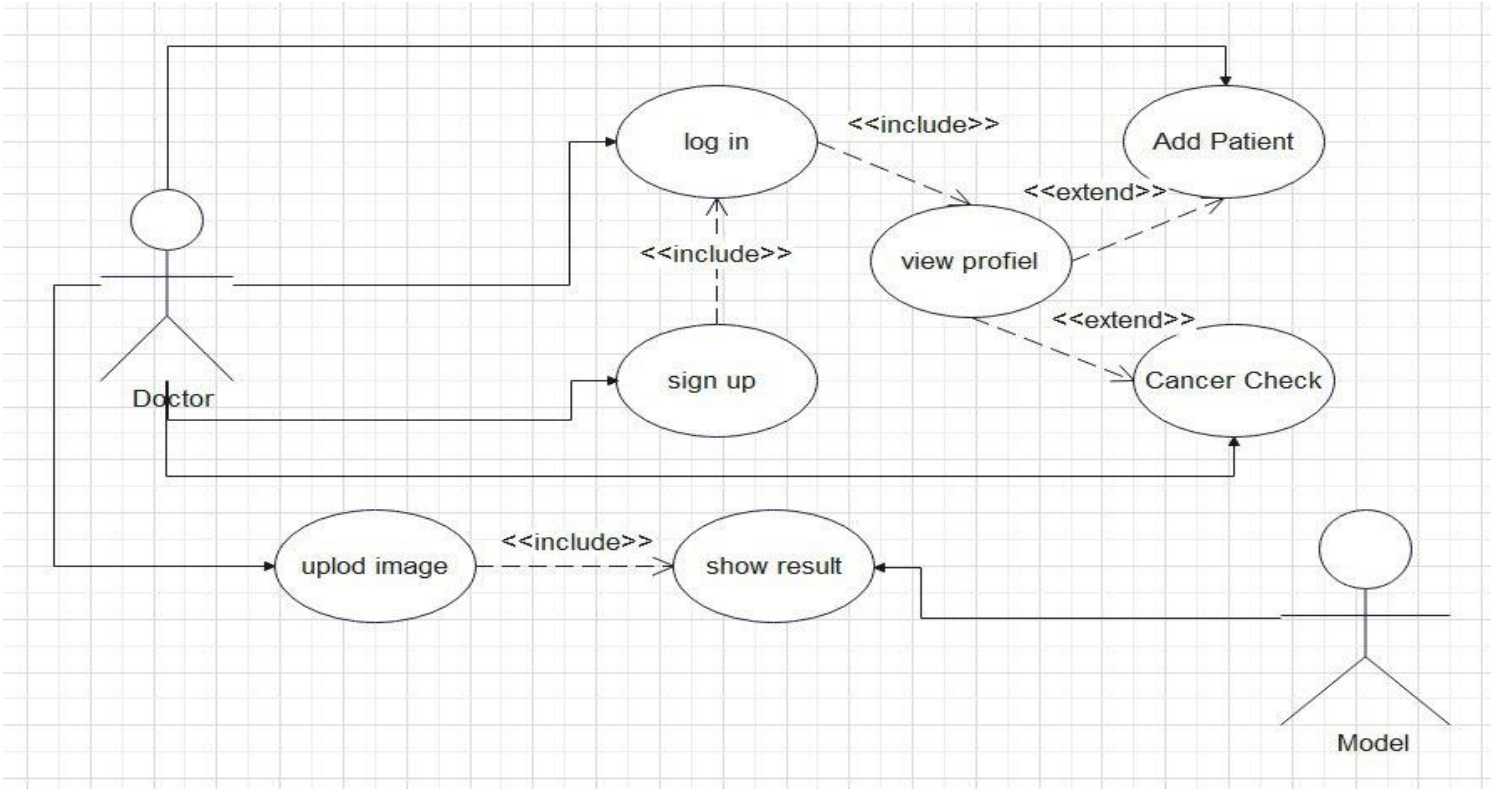


Fig 3.1

Chapter Four
System Design

System Component Diagram

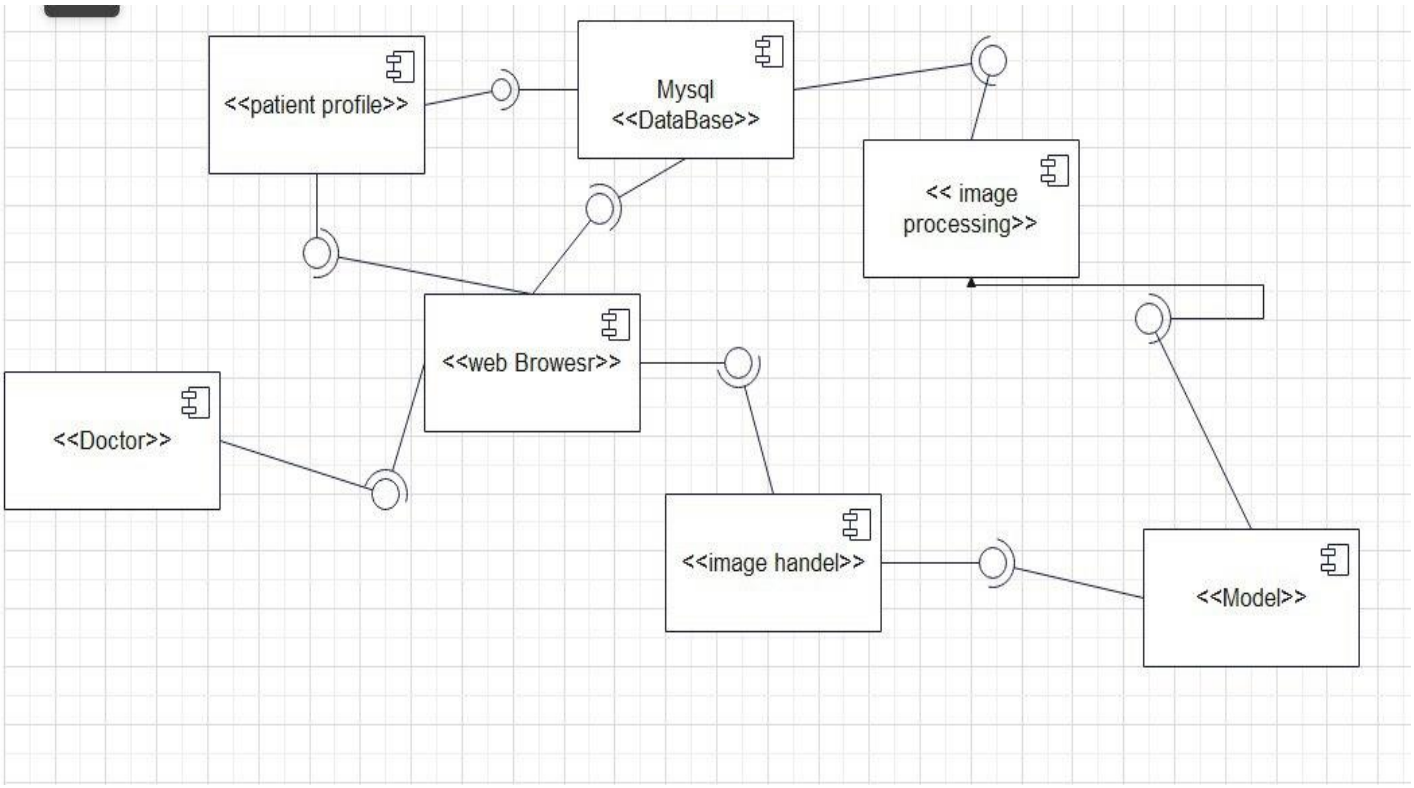


Fig 4.1

System class Diagram

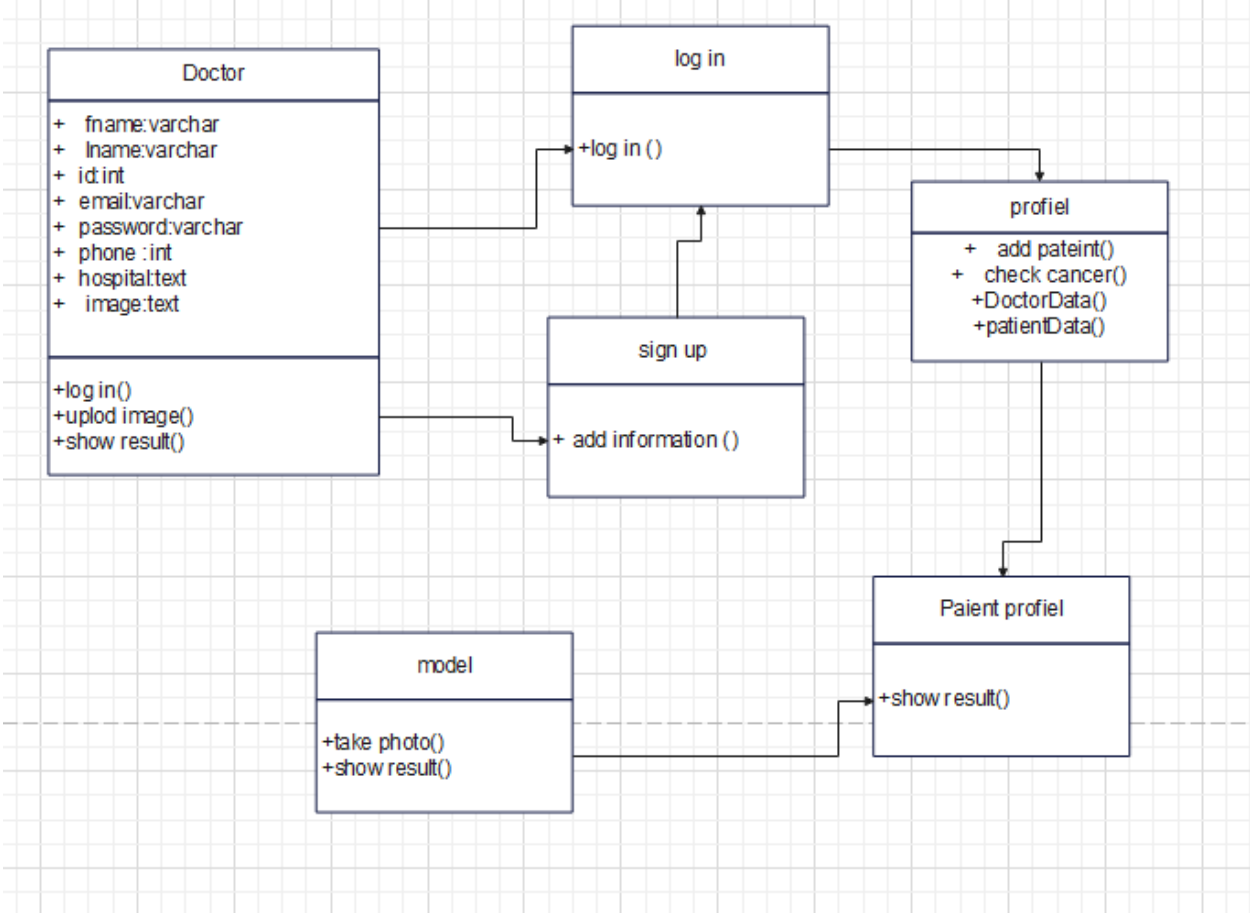


Fig 4.2

Sequence Diagrams

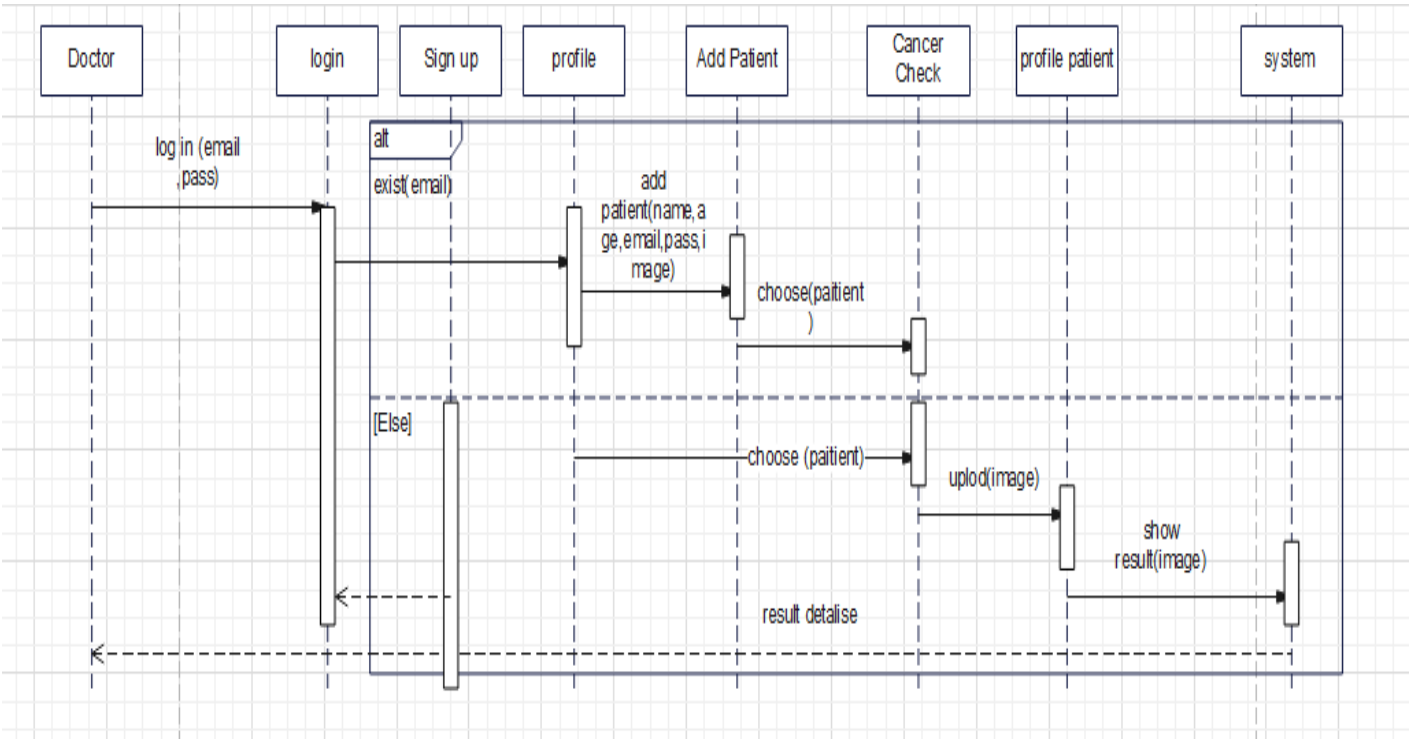
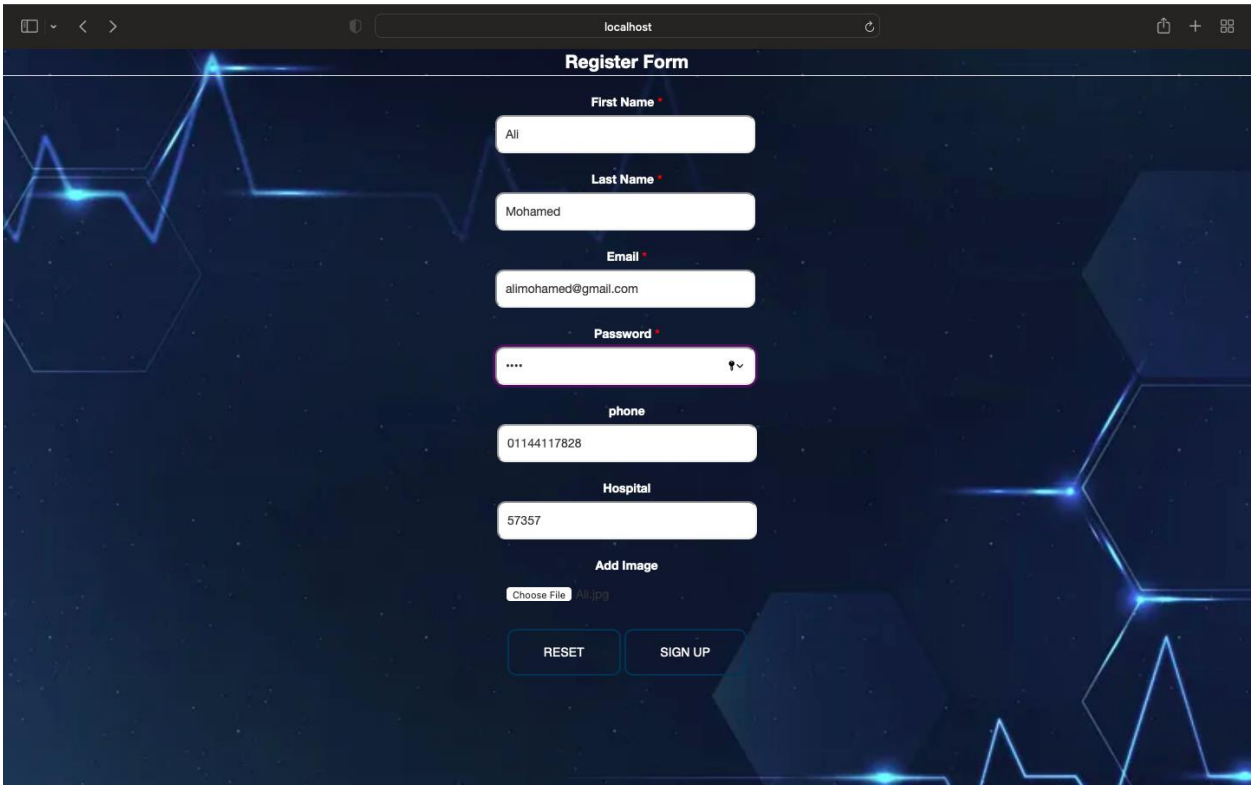


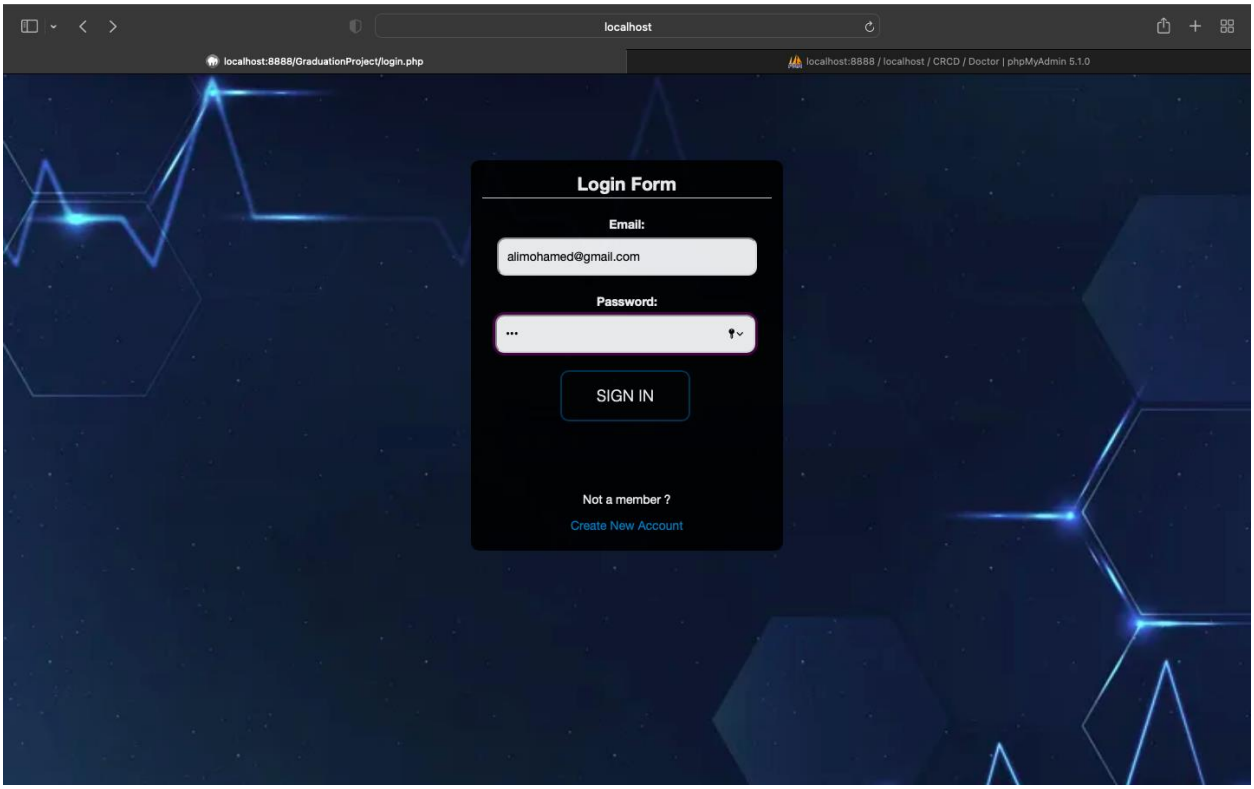
Fig 4.3

Chapter Five
Implementation and Testing

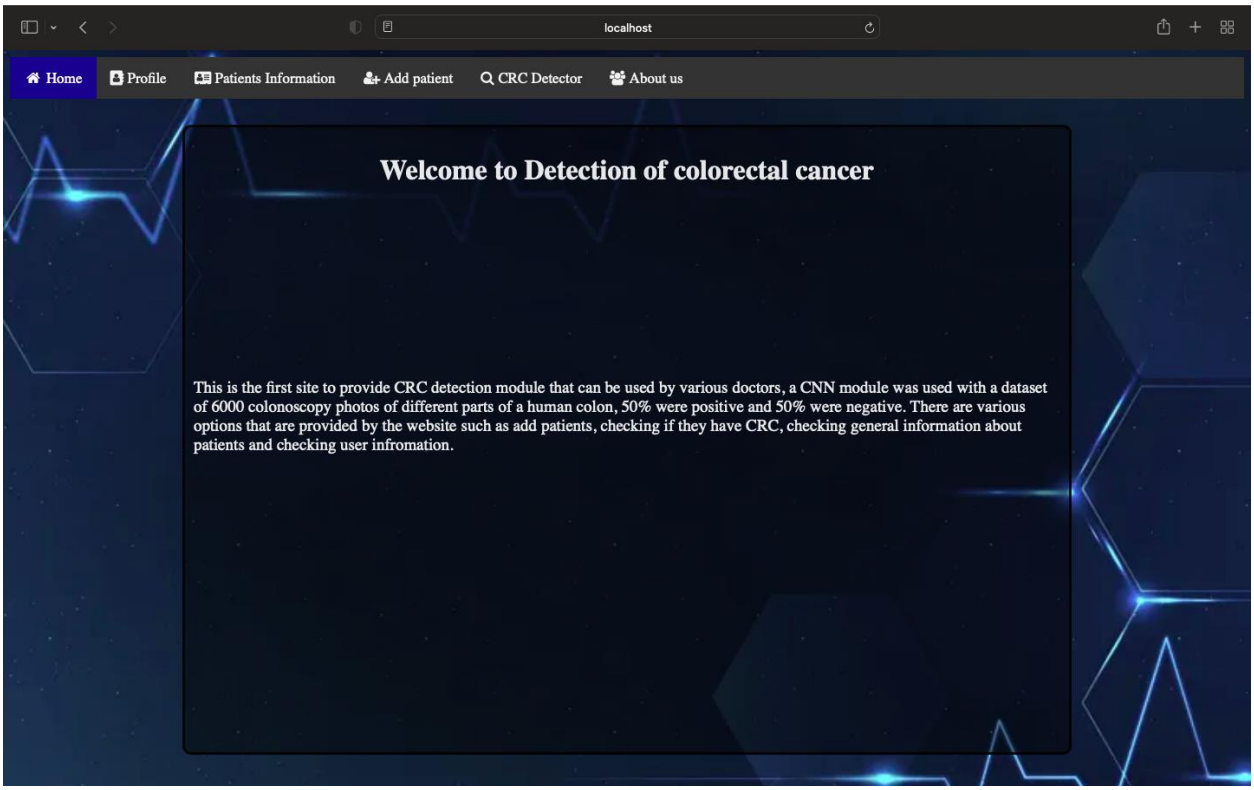
Register Form



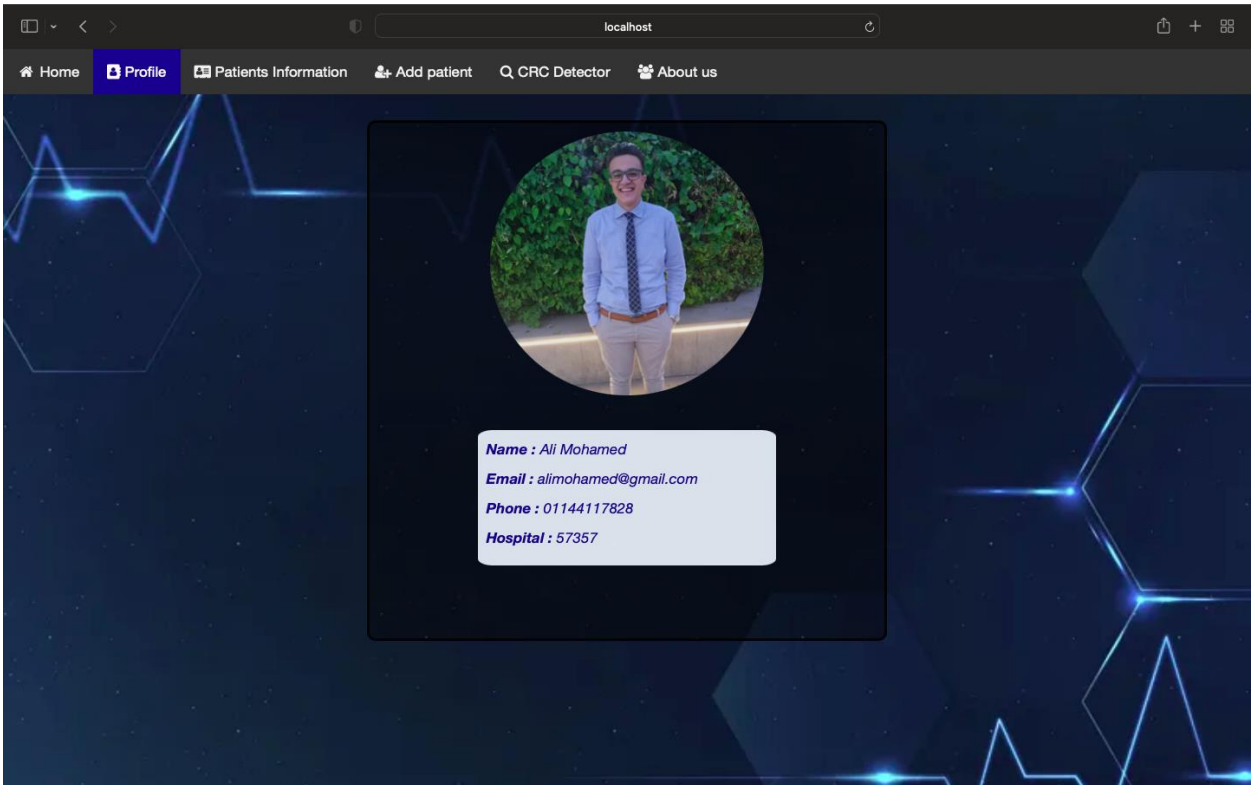
Log in Page


localhost
localhost:8888/GraduationProject/login.php
localhost:8888 / localhost / CRCD / Doctor | phpMyAdmin 5.1.0
Login Form
Email:
alimohamed@gmail.com
Password:
SIGN IN
Not a member ?
Create New Account

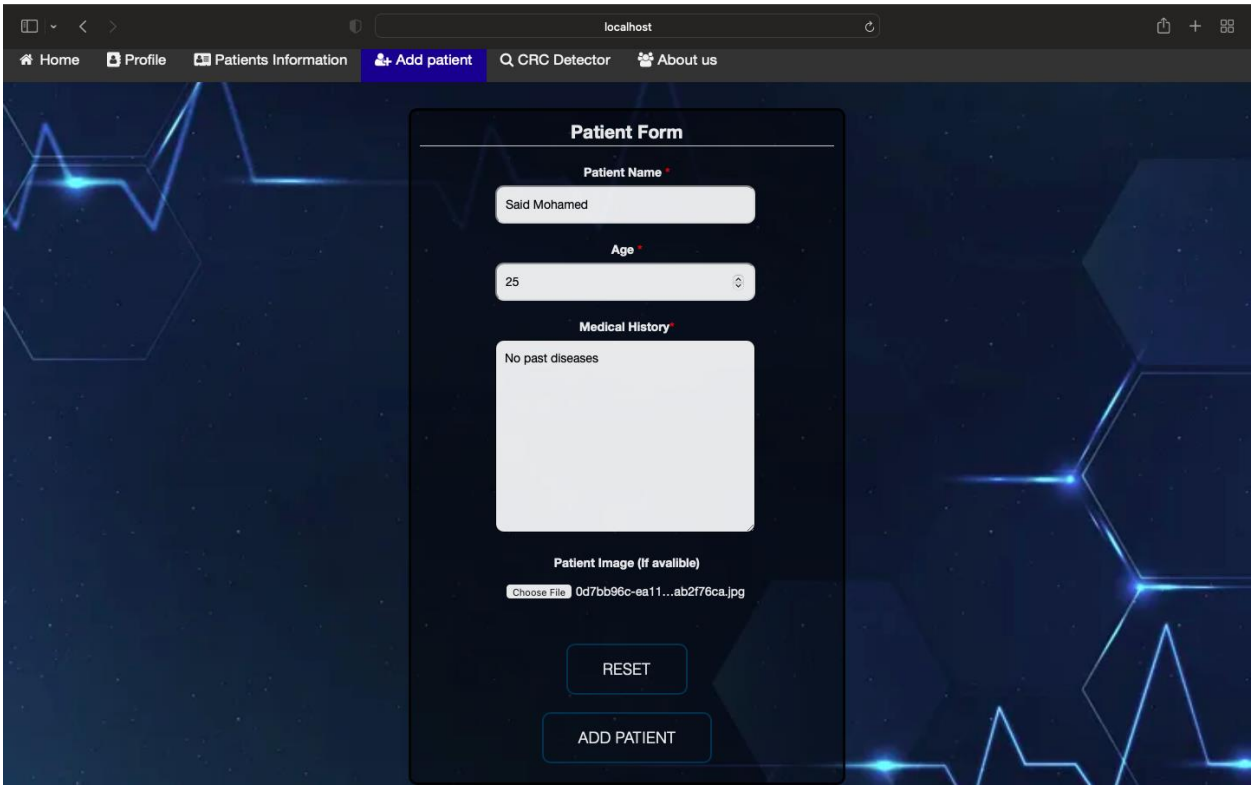
Home

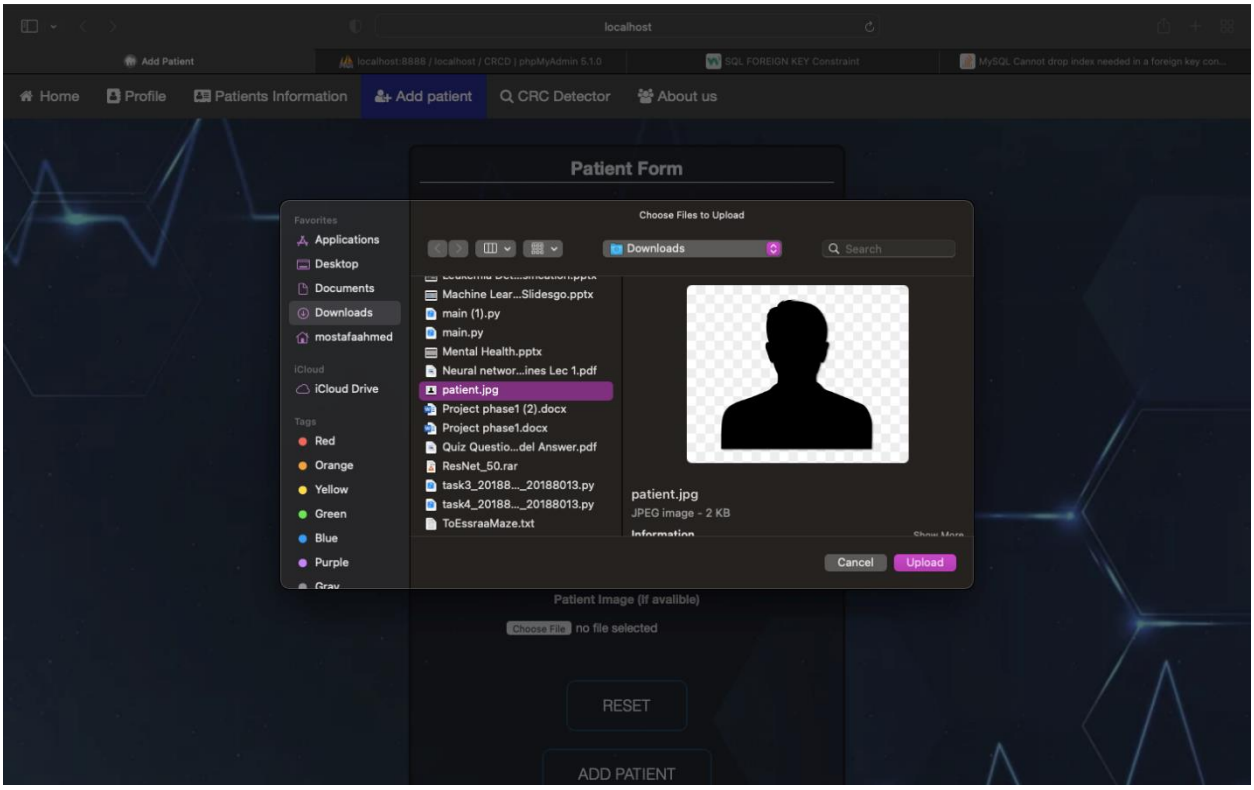

localhost
Home
Profile
Patients Information
Add patient
CRC Detector
About us
Welcome to Detection of colorectal cancer
This is the first site to provide CRC detection module that can be used by various doctors, a CNN module was used with a dataset of 6000 colonoscopy photos of different parts of a human colon, 50% were positive and 50% were negative. There are various options that are provided by the website such as add patients, checking if they have CRC, checking general information about patients and checking user infromation.

Profile

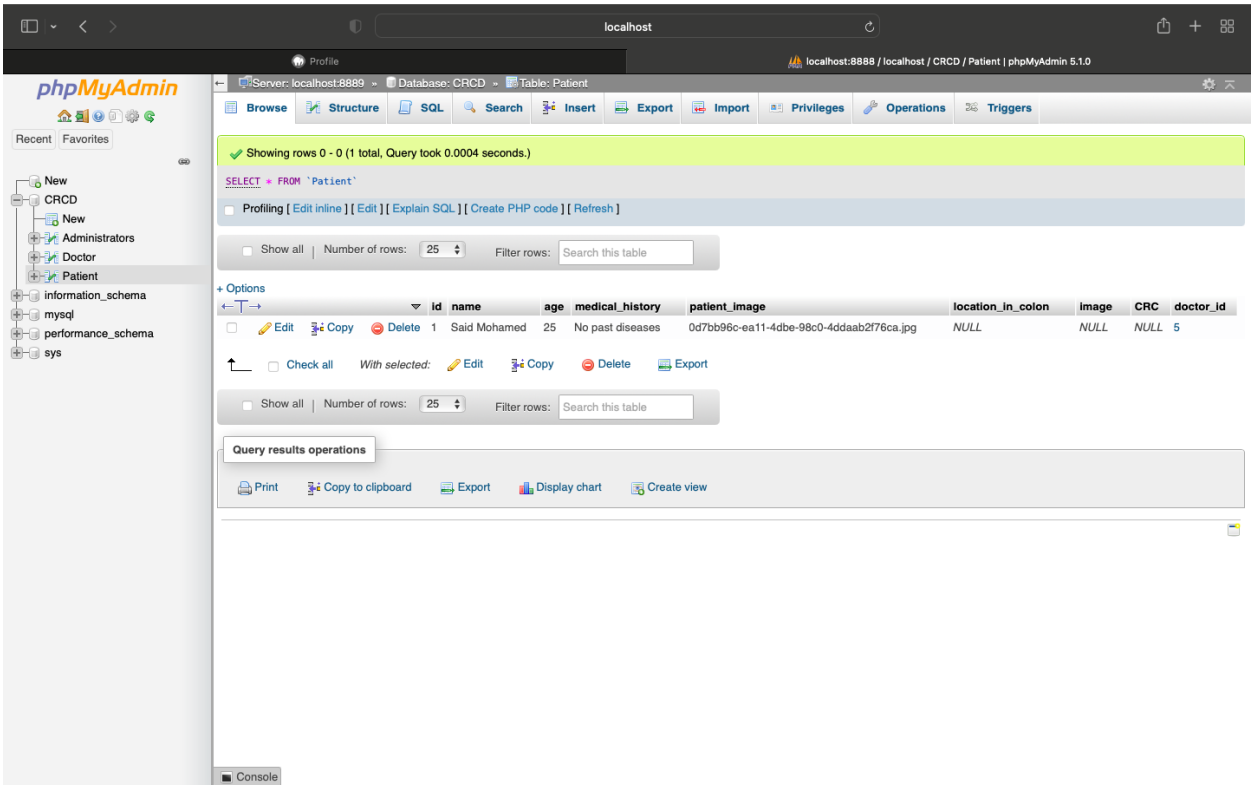

localhost
Home
Profile
Patients Information
Add patient
CRC Detector
About us
Name : Ali Mohamed
Email : alimohamed@gmail.com
Phone : 01144117828
Hospital : 57357

Add Patient


localhost
Home
Profile
Patients Information
Add patient
CRC Detector
About us
Patient Form
Patient Name *
Said Mohamed
Age *
25
Medical History*
No past diseases
Patient Image (If avalible)
Choose File
0d7bb96c-ea11...ab2f76ca.jpg
RESET
ADD PATIENT

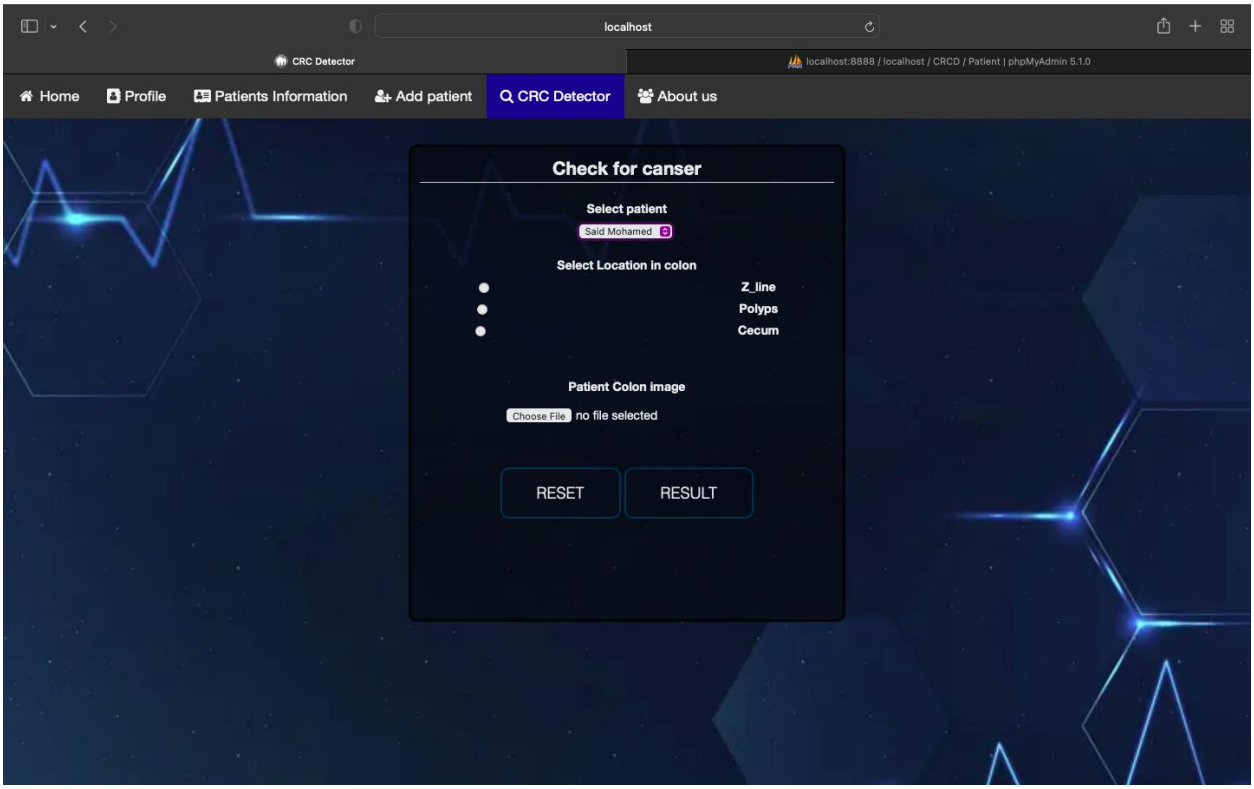

localhost
Add Patient
Home
Profile
Patients Information
Add patient
CRC Detector
About us
Patient Form
Choose Files to Upload
Favorites
Applications
Desktop
Documents
Downloads
mostafaahmed
iCloud
iCloud Drive
Tags
Red
Orange
Yellow
Green
Blue
Purple
Downloads
Search
Machine Lear...Slidesgo.pptx
main (1).py
main.py
Mental Health.pptx
Neural networ...ines Lec 1.pdf
patient.jpg
Project phase1 (2).docx
Project phase1.docx
Quiz Questio...del Answer.pdf
ResNet_50.rar
task3_20188..._20188013.py
task4_20188..._20188013.py
ToEssraaMaze.txt
patient.jpg
JPEG image - 2 KB
Cancel
Upload
Patient Image (If avalible)
Choose File
no file selected
RESET
ADD PATIENT

Data Base inserted in Patient table

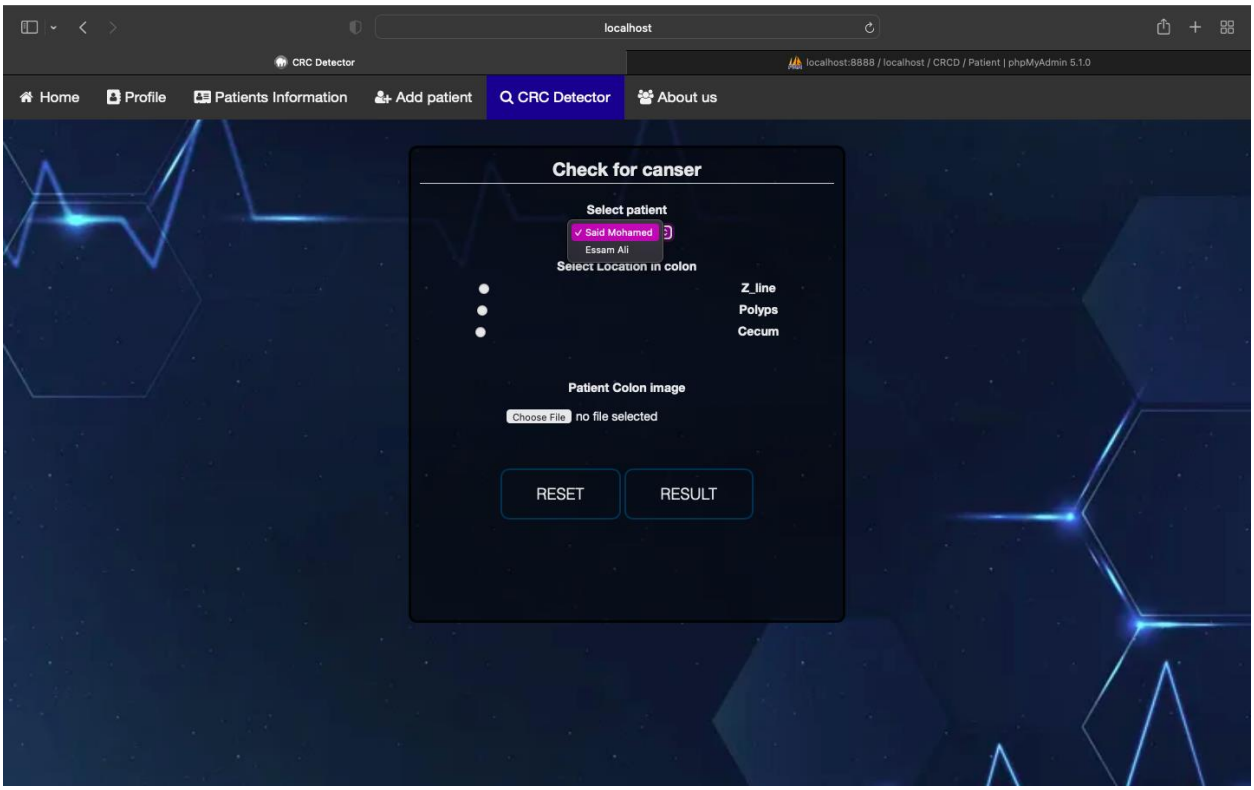


CRC Detector

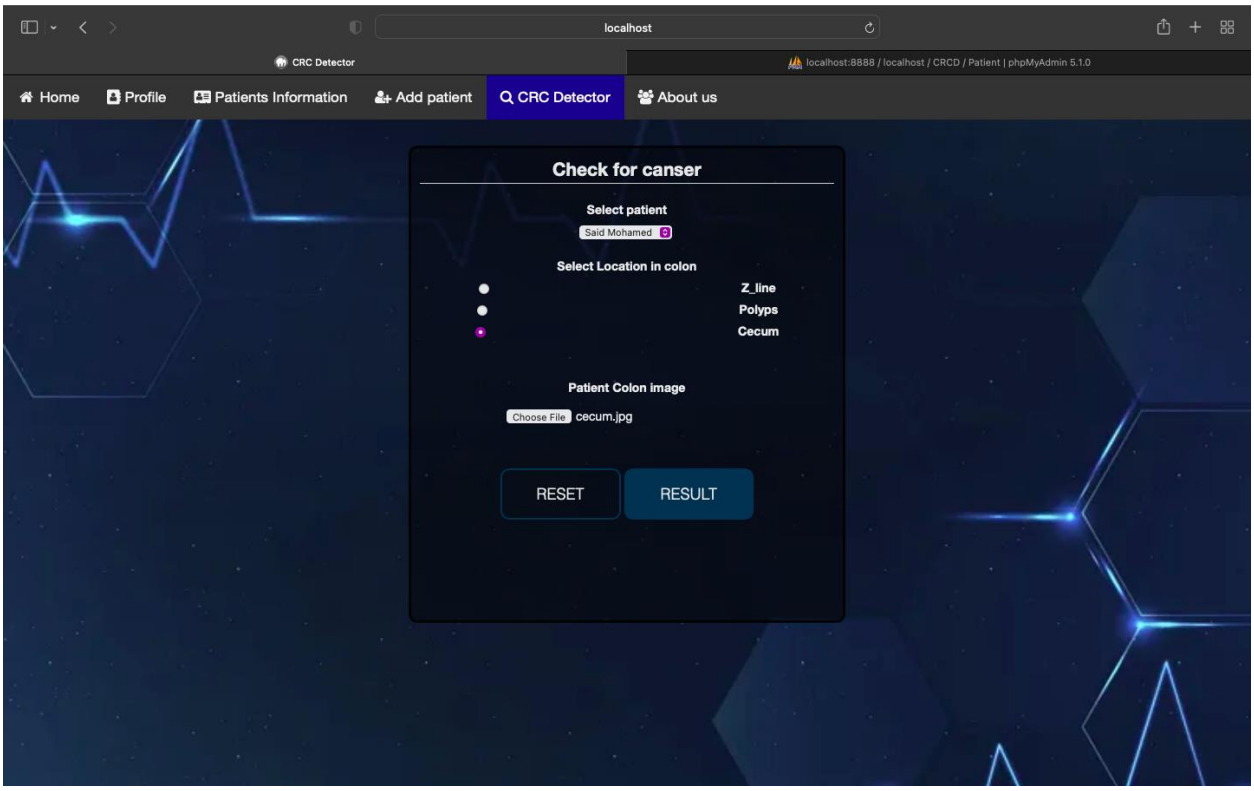
1.



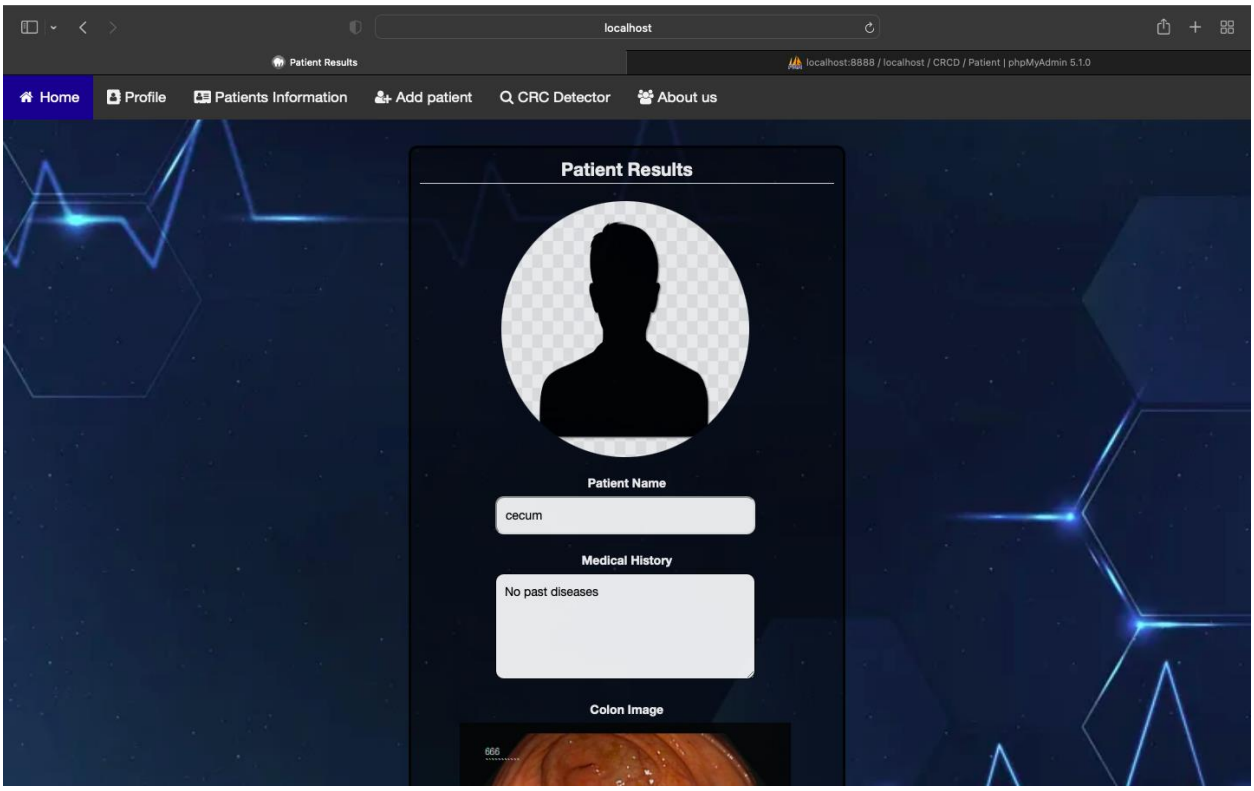
2.

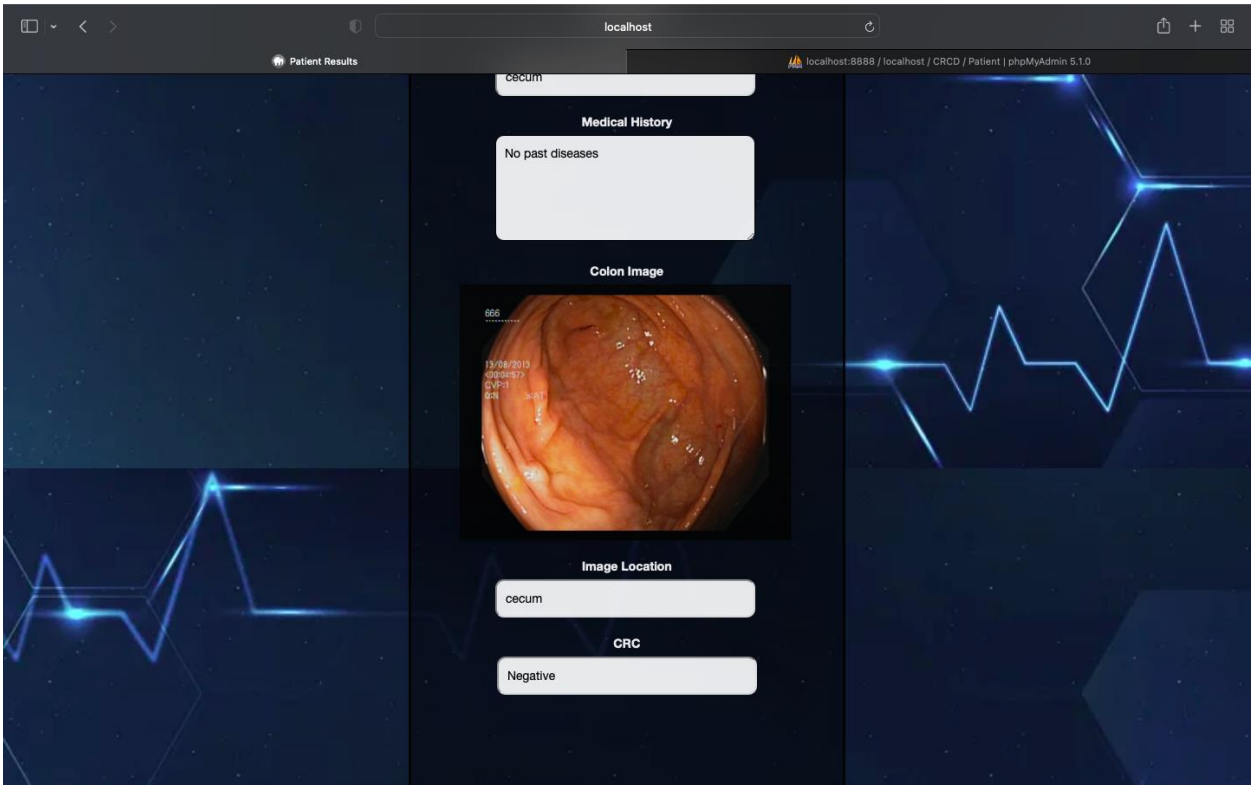


3.

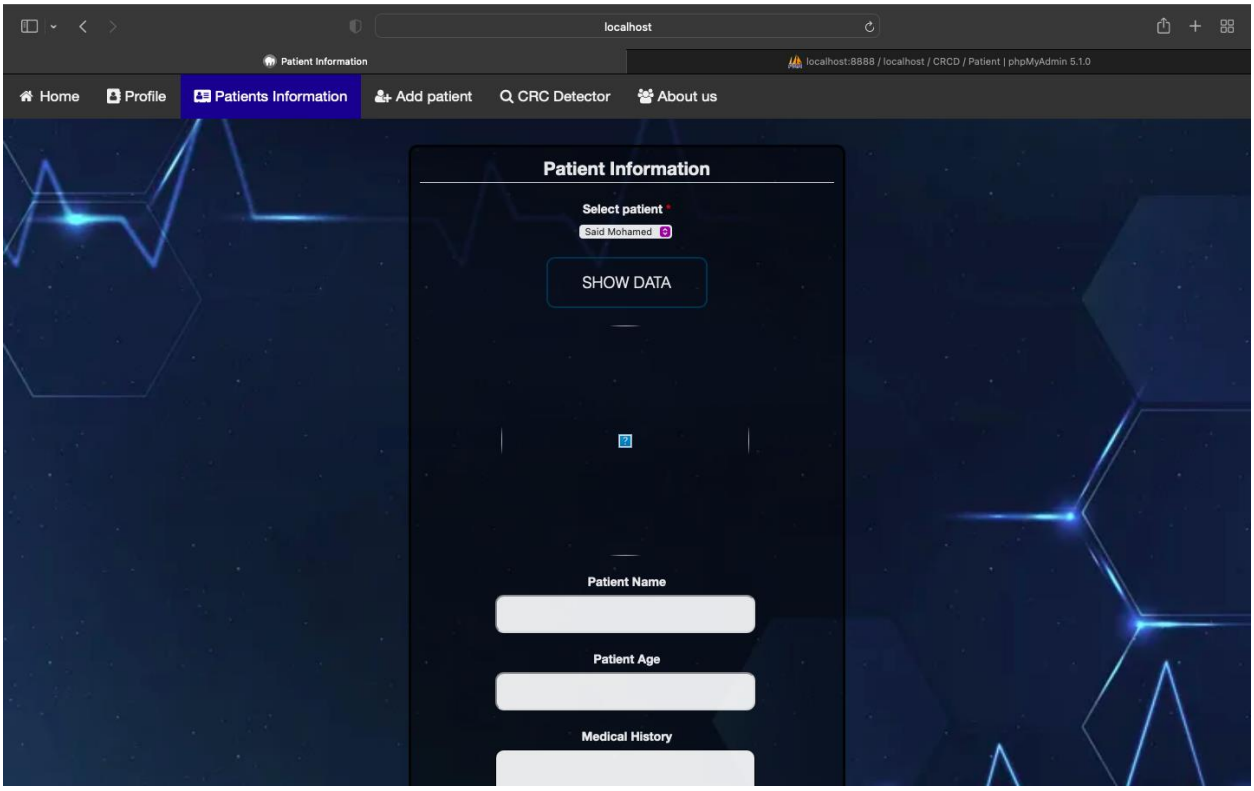

localhost
CRC Detector
localhost:8888 / localhost / CRCD / Patient | phpMyAdmin 5.1.0
Home
Profile
Patients Information
Add patient
CRC Detector
About us
Check for canser
Select patient
Said Mohamed
Select Location in colon
Z_line
Polyps
Cecum
Patient Colon image
Choose File
cecum.jpg
RESET
RESULT

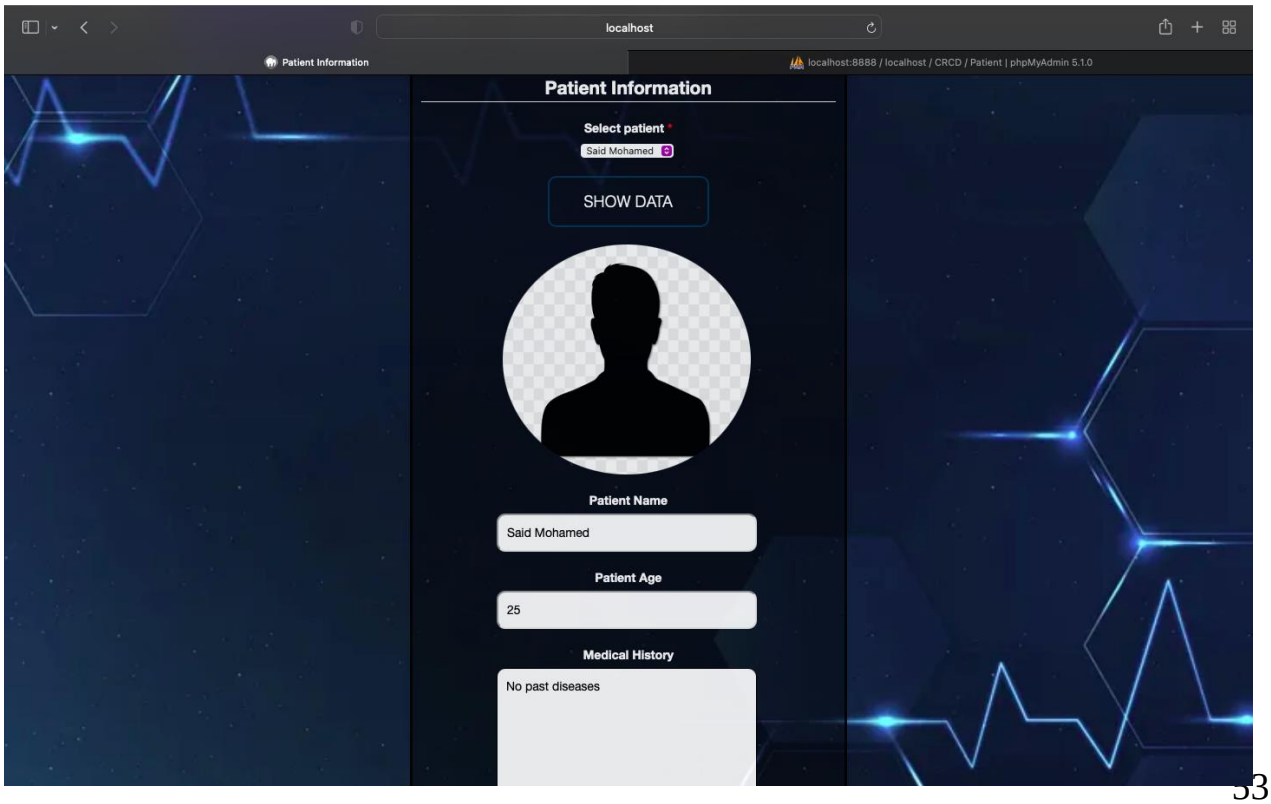
Patient Result

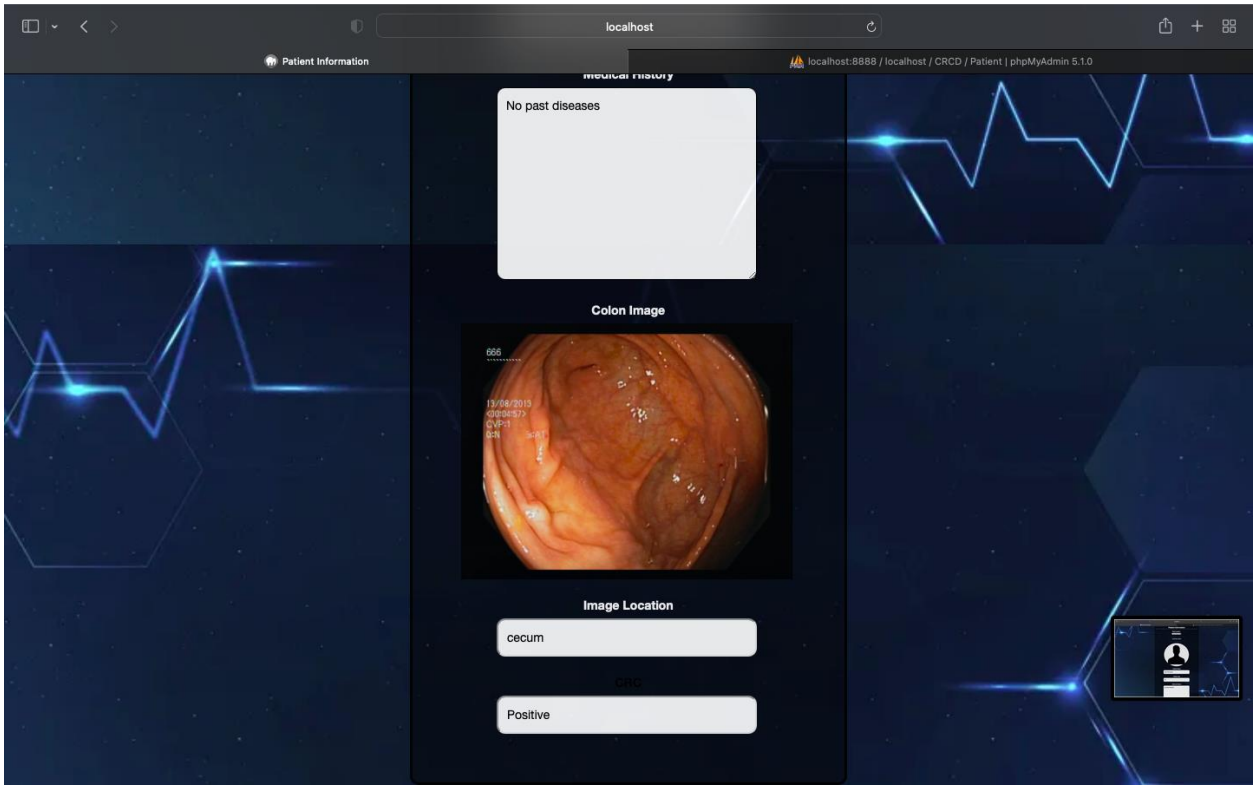




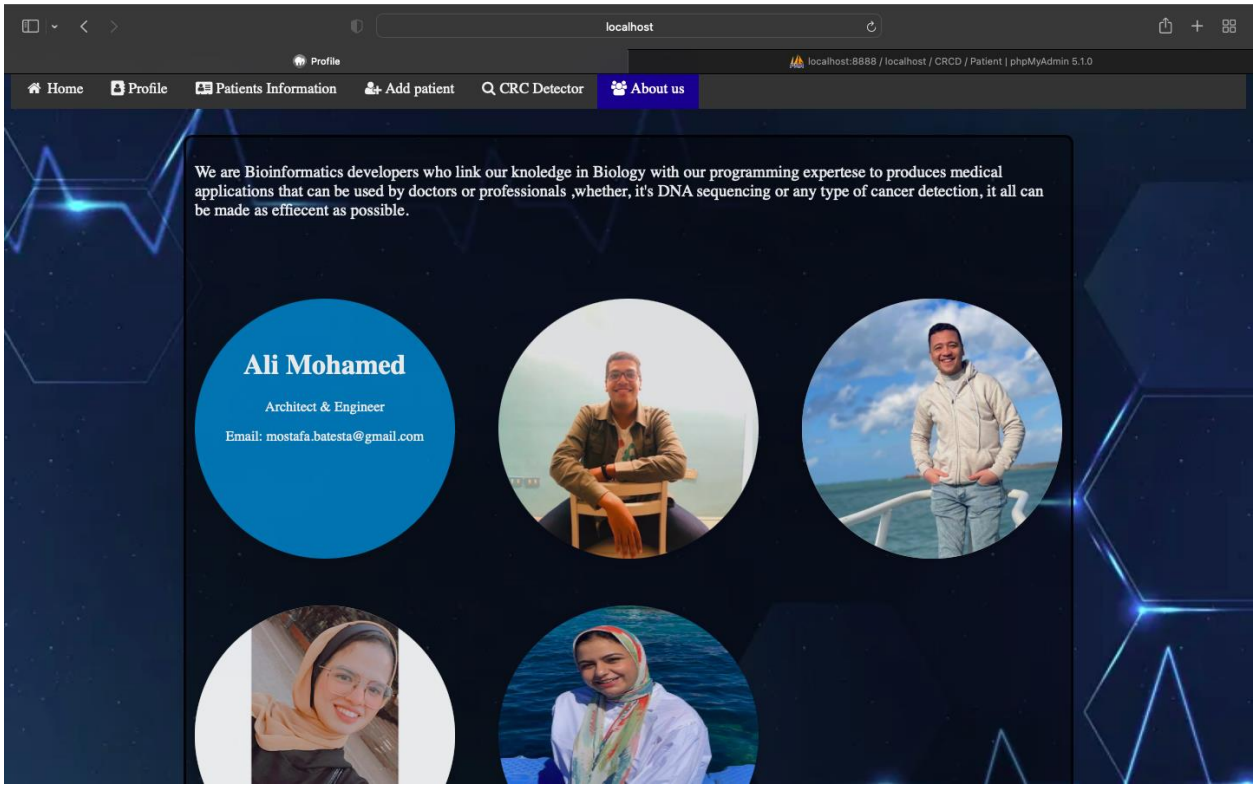
Patient Information

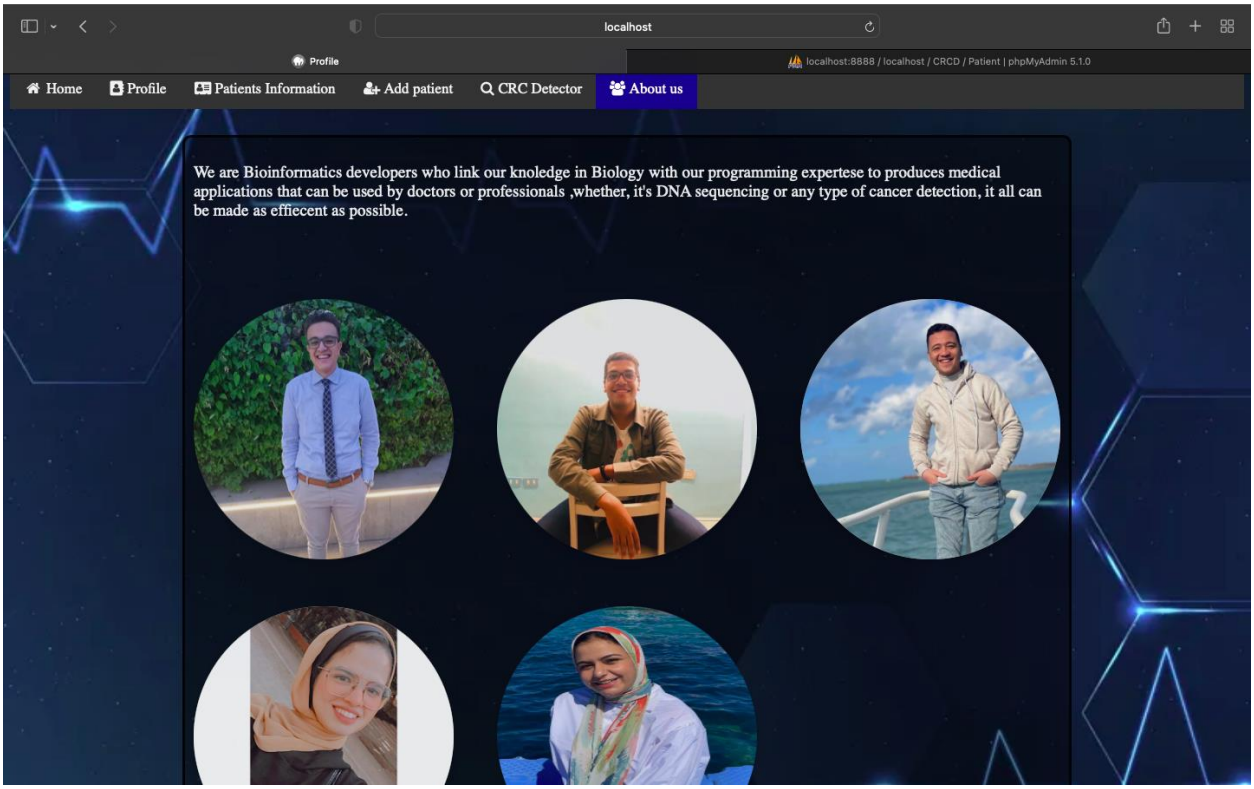




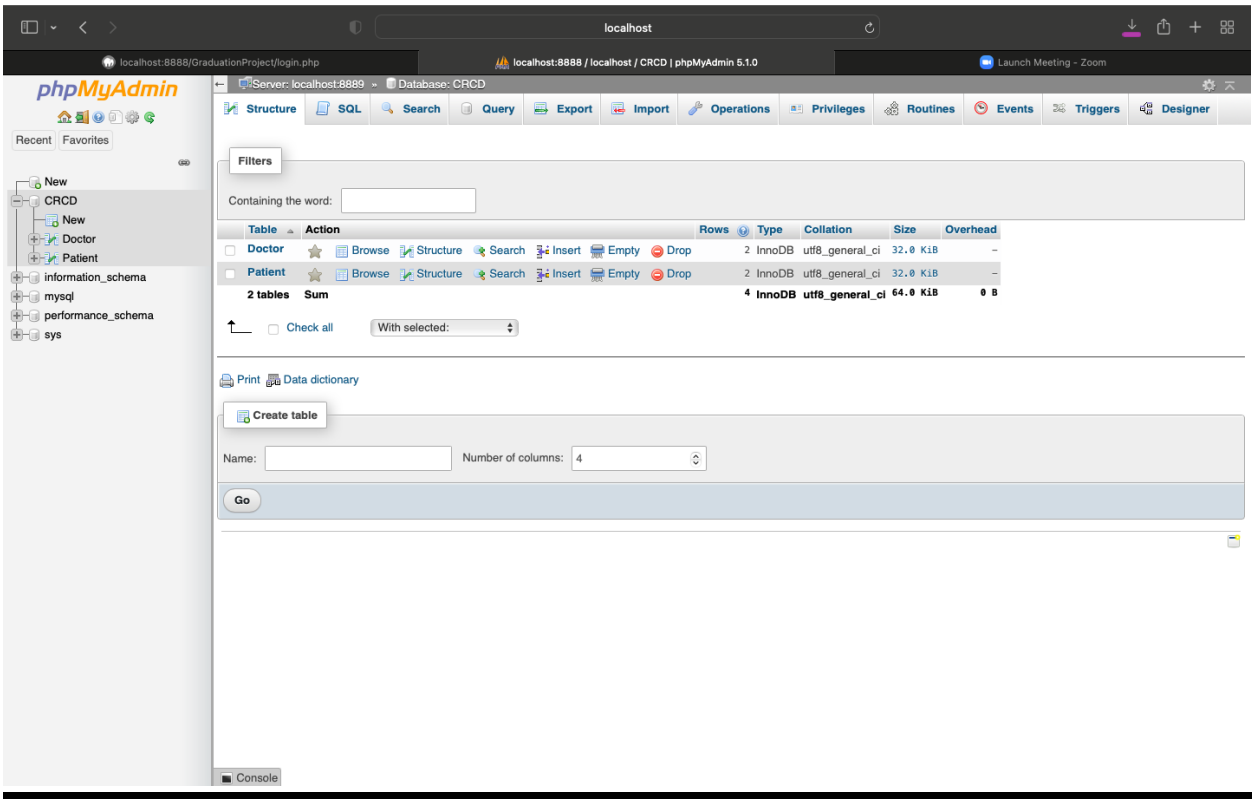

localhost
Patient Information
localhost:8888 / localhost / CRCD / Patient | phpMyAdmin 5.1.0
No past diseases
Colon Image
Image Location
cecum
CRC
Positive

About Us

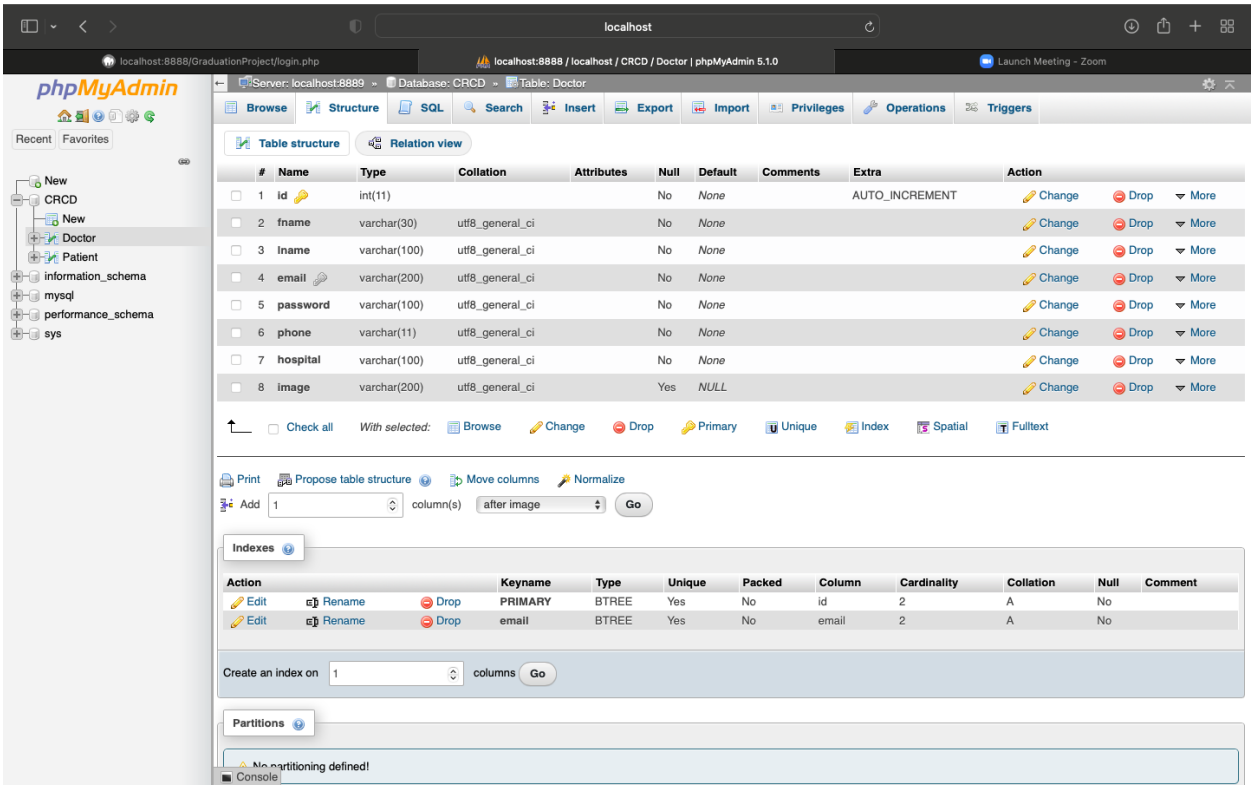

localhost
Profile
localhost:8888 / localhost / CRCD / Patient | phpMyAdmin 5.1.0
Home
Profile
Patients Information
Add patient
CRC Detector
About us
We are Bioinformatics developers who link our knoledge in Biology with our programming expertese to produces medical applications that can be used by doctors or professionals ,whether, it's DNA sequencing or any type of cancer detection, it all can be made as effiecent as possible.
Ali Mohamed
Architect & Engineer
Email: mostafa.batesta@gmail.com

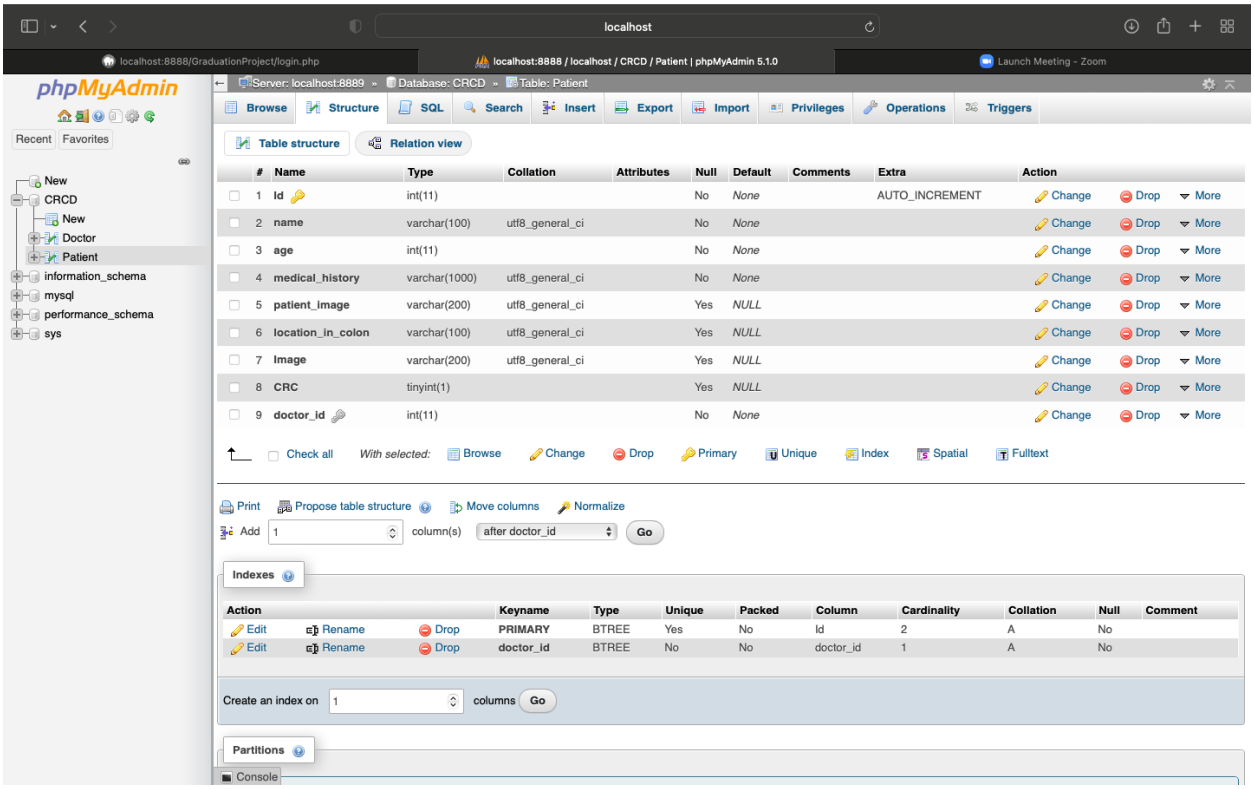

localhost
Profile
localhost:8888 / localhost / CRCD / Patient | phpMyAdmin 5.1.0
Home
Profile
Patients Information
Add patient
CRC Detector
About us
We are Bioinformatics developers who link our knoledge in Biology with our programming expertese to produces medical applications that can be used by doctors or professionals ,whether, it's DNA sequencing or any type of cancer detection, it all can be made as effiecent as possible.

Database Schema



Tables


localhost
localhost:8888/GraduationProject/login.php
localhost:8888 / localhost / CRCD / Doctor | phpMyAdmin 5.1.0
Launch Meeting - Zoom
phpMyAdmin
Recent Favorites
New
CRCD
New
Doctor
Patient
information_schema
mysql
performance_schema
sys
Server: localhost:8889 » Database: CRCD » Table: Doctor
Browse Structure SQL Search Insert Export Import Privileges Operations Triggers
Table structure Relation view
Name Type Collation Attributes Null Default Comments Extra Action
1 id int(11) No None AUTO_INCREMENT Change Drop More
2 fname varchar(30) utf8_general_ci No None Change Drop More
3 lname varchar(100) utf8_general_ci No None Change Drop More
4 email varchar(200) utf8_general_ci No None Change Drop More
5 password varchar(100) utf8_general_ci No None Change Drop More
6 phone varchar(11) utf8_general_ci No None Change Drop More
7 hospital varchar(100) utf8_general_ci No None Change Drop More
8 image varchar(200) utf8_general_ci Yes NULL Change Drop More
Check all With selected: Browse Change Drop Primary Unique Index Spatial Fulltext
Print Propose table structure Move columns Normalize
Add 1 column(s) after image Go
Indexes
Action Keyname Type Unique Packed Column Cardinality Collation Null Comment
Edit Rename Drop PRIMARY BTREE Yes No id 2 A No
Edit Rename Drop email BTREE Yes No email 2 A No
Create an index on 1 columns Go
Partitions
No partitioning defined!
Console


localhost
localhost:8888/GraduationProject/login.php
localhost:8888 / localhost / CRCD / Patient | phpMyAdmin 5.1.0
Launch Meeting - Zoom
phpMyAdmin
Recent Favorites
New
CRCD
New
Doctor
Patient
information_schema
mysql
performance_schema
sys
Server: localhost:8889 » Database: CRCD » Table: Patient
Browse Structure SQL Search Insert Export Import Privileges Operations Triggers
Table structure Relation view
Name Type Collation Attributes Null Default Comments Extra Action
1 Id int(11) No None AUTO_INCREMENT Change Drop More
2 name varchar(100) utf8_general_ci No None Change Drop More
3 age int(11) No None Change Drop More
4 medical_history varchar(1000) utf8_general_ci No None Change Drop More
5 patient_image varchar(200) utf8_general_ci Yes NULL Change Drop More
6 location_in_colon varchar(100) utf8_general_ci Yes NULL Change Drop More
7 Image varchar(200) utf8_general_ci Yes NULL Change Drop More
8 CRC tinyint(1) Yes NULL Change Drop More
9 doctor_id int(11) No None Change Drop More
Check all With selected: Browse Change Drop Primary Unique Index Spatial Fulltext
Print Propose table structure Move columns Normalize
Add 1 column(s) after doctor_id Go
Indexes
Action Keyname Type Unique Packed Column Cardinality Collation Null Comment
Edit Rename Drop PRIMARY BTREE Yes No Id 2 A No
Edit Rename Drop doctor_id BTREE No No doctor_id 1 A No
Create an index on 1 columns Go
Partitions
Console

Appendix

Phase one model using supervised technique (SVM) with 87% accuracy

```
def svm(svm_x_train, svm_y_train, svm_x_test, svm_y_test):
    svc_model = SVC(C=.1, kernel='poly', gamma=1)
    svc_model.fit(svm_x_train, svm_y_train)
    prediction = svc_model.predict(svm_x_test)

    filename = 'finalized_model.sav'
    pickle.dump(svc_model, open(filename, 'wb'))
    return accuracy_score(svm_y_test, prediction) * 100
```

Phase two models using unsupervised technique (CNN)

First model with 73%accuracy

```
KerasModel = keras.models.Sequential([
    keras.layers.Conv2D(32, kernel_size=(3, 3), input_shape=(size, size, 1)),

    keras.layers.MaxPool2D(2, 2),

    keras.layers.Conv2D(64, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2, 2),

    keras.layers.Conv2D(128, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2, 2),

    keras.layers.Conv2D(256, kernel_size=(5, 5)),
    keras.layers.MaxPool2D(2, 2),

    keras.layers.Conv2D(512, kernel_size=(5, 5)),
    keras.layers.MaxPool2D(2, 2),

    keras.layers.Flatten(),

    keras.layers.Dense(6, activation='softmax')
])
```

Second model with 88%accuracy

```
KerasModel = keras.models.Sequential([
    keras.layers.Conv2D(200, kernel_size=(3, 3),  activation='relu',input_shape=(size, size, 3)),
    keras.layers.MaxPool2D(2,2),
    keras.layers.Dropout(rate=0.2),


    keras.layers.Conv2D(240, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2,2),
    keras.layers.Dropout(rate=0.2),

    keras.layers.Conv2D(280, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2, 2),
    keras.layers.Dropout(rate=0.2),


    keras.layers.Conv2D(320, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2,2),
    keras.layers.Dropout(rate=0.2),

    keras.layers.Conv2D(360, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2,2),
    keras.layers.Dropout(rate=0.2),
```

```
    keras.layers.Conv2D(400, kernel_size=(3, 3)),
    keras.layers.MaxPool2D(2,2),
    keras.layers.Dropout(rate=0.2),


    keras.layers.Flatten(),
    keras.layers.Dense(512, activation='relu'),
    keras.layers.Dense(512, activation='relu'),
    keras.layers.Dense(512, activation='relu'),

    keras.layers.Dropout(rate=0.2),

    keras.layers.Dense(6, activation='softmax',kernel_regularizer=l2(0.01),
                       bias_regularizer=l2(0.01))
])
```

Third model CNN VGG19 with 87%accuracy

```
inputs = Input(shape=(224,224,3))
base_model = VGG19(include_top=False, weights='imagenet', input_shape=(224, 224,3))
base_model.trainable = False

added_layers = base_model(inputs, training=False)
added_layers = Flatten()(added_layers)
added_layers = Dense(100,activation='relu')(added_layers)
added_layers = Dense(50, activation='relu')(added_layers)
final_layer = Dense(6, activation='softmax')(added_layers)

KerasModel = Model(inputs, final_layer)
KerasModel.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
epochs = 25
KerasModel.fit(x_train, to_categorical(y_train), epochs=epochs, batch_size=64, verbose=1,
               validation_data=(x_valid, to_categorical(y_valid)))
```

Fourth model CNN ResNet50 with 93%accuracy

```
inputs = Input(shape=(224,224,3))
base_model = ResNet50(include_top=False, weights='imagenet', input_shape=(224, 224,3))
base_model.trainable = False

added_layers = base_model(inputs, training=False)
added_layers = Flatten()(added_layers)
added_layers = Dense(100,activation='relu')(added_layers)
added_layers = Dense(50, activation='relu')(added_layers)
final_layer = Dense(6, activation='softmax')(added_layers)

KerasModel = Model(inputs, final_layer)
KerasModel.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
epochs = 25
KerasModel.fit(x_train, to_categorical(y_train), epochs=epochs, batch_size=64, verbose=1,
               validation_data=(x_valid, to_categorical(y_valid)))
```

Poster3

Early detection of colorectal cancer using machine learning & deep learning

Project logo

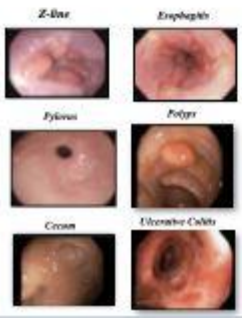
Mostafa, Ali, Anas, Esraa, and Rawan
Supervisor: Dr. Hanaa & Ahmed

Abstract

Colorectal cancer (CRC) begins in the colon or the rectum. These diseases can likewise be called colon malignant growth or rectal cancer, contingent upon where they start. Colon malignant growth and rectal disease are regularly assembled in light of the fact that they share many highlights practically speaking. Colorectal disease is second reason for death in many communities additionally early conclusion have a more prominent possibility of survival .So colonoscopy represents a very important diagnostic modality for screening for colorectal cancer, we address a current issue in medical picture handling, the discovery of colorectal disease from colonoscopy videos. As per overall malignant growth measurements, colorectal disease is perhaps the most well-known disease. The most common way of screening and the expulsion of pre-malignant cells from the digestive organ is an essential undertaking to date. The traditional manual process is dependent on the expertise of the medical practitioner. We have Two phases in this project The first stage we utilized SVM model with an accuracy of 87 % in second stage we utilized deep learning CNN models with some accuracies.so, The main objective of this project is early detection of colon cancer based on medical image processing extracted from colonoscopy videos.

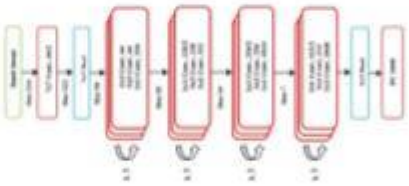
Introduction

Cancer is Billions of cells wear out or get damaged. Each time a cell is destroyed, the body makes an entirely new cell to replace it, in an effort to create a cell that is a perfect copy of the cell that was destroyed
CRC starts when the process of the normal replacement colon lining cells deviate. Health care professionals are certain that colorectal cancer is not contagious (a person cannot catch the disease from a cancer patient). Some people are more likely to develop colorectal cancer than others
Symptoms is (Blood in stool ,Abdominal pain, Fatigue ,unexplained)
Factors is (Age, Races, Diet and colorectal cancer, Inherited syndromes, A sedentary lifestyle, Diabetes, Obesity, Smoking, Alcohol, Radiation therapy for cancer)
there is 4 stages of CRC



Methods

Dataset contain 6000 images from three anatomical landmark each landmark has 1000 for normal case and 1000 for cancer case
Removing Nonrelated shape from images
Blurring the images
Drawing contours on RGB images
Flatten images
Split the data into train and validate and test with Shuffle and it was used to build the model
Classifying images using different classifiers
Phase one model using supervised technique (SVM) with 87% accuracy
Phase two models using different architectures of CNN the highest accuracy is 95% with RESNET50



Primarily Design

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.

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.

- Abdullah M, Rani AA, Simadibrata M, Fauzi A, Syam AF (2012) The value of fecal tumor M2 pyruvate kinase as a diagnostic tool for colorectal cancer screening. Acta med Indones 44:94–99.