Chapter 5.
System Description and Evaluation Models

5.1 Colonoscopy Polyp Classification and Detection

Early detection plays a significant role in fighting CRC. It not only brings down the mortality but also prevents excessive treatment cost by diagnosing before CRC spreads to distant organs. According to research ***[1]***, the stages at which the disease is diagnosed highly correlate to survival, with a **90% 5-year survival rate** for the localized stage, 70% for the regional stage, and 10% for distant metastatic cancer. Another reason we should rely on early detection is due to the nature of the symptoms and development of CRC. Although no symptoms can be easily observed before the tumor reaches a certain size according to research ***[2]***, it would typically take several years to as long as a decade for CRC to develop, starting from precancerous polyps. Both facts add up to show the significance and potential of diagnosing CRC by regular screening at an early stage, even before polyps become cancerous.

Colonoscopy: is the recommended CRC visual examination screening method. The advantages of colonoscopy include high sensitivity, ability to remove lesions at detection and full access to proximal and distal portions of the colon. The colonoscopy can reach a sensitivity of 95% in detecting CRC according to Rex *et al.* ***[3]***. The disadvantages are mostly related to the way colonoscopy is conducted. At least one day before the test, it requires a complicated bowel preparation, which requires the participant to change diet and take medicine to cause diarrhea. During the test, sedation or anesthesia might be performed, and there is a risk of post colonoscopy bleeding. Thus, the suggested 10-year screening interval has a low compliance rate. Narrow-Band Imaging (NBI) is a newly developed technique by modifying light sources using optical filters in an endoscopy system ***[4]***. Compared to normal colonoscopy, intensified lights of a certain wavelength can better present the mucosal morphology and vascular pattern. Studies show that NBI performs better in CRC detection than conventional colonoscopy ***[5]***.

Colonoscopy has obvious advantages over its alternatives. However, its performance depends on several variables, like the bowel preparation, the number of polyps, and the part of the colon where the polyps are located. Furthermore, human-factors can influence the screening sensitivity and specificity. Inexperienced gastroenterologists have higher miss detection rates compared to those who are well-trained. According to *Leufkens et al.* ***[6]***, participants before training showed significantly lower performance than post-training results. Colonoscopy is also subjected to the physical and mental fatigue of the gastroenterologists. The screening process requires prolonged concentration and is usually repeated throughout the day. A study by Chan *et al.* ***[7]*** showed that 20% more polyps are detected from early morning screenings.

It is obvious that a fine-grained deep learning framework to automatically detect polyps is needed to help physicians locate and classify the lesions. This deep learning framework can assist physicians during screening in real-time and prompt the detected region and polyp category. Thus, such a computer-aided system can help eliminate the miss rate due to physical and mental fatigue and allow the gastroenterologists to focus on regions where lesions exist. This automated system also ensures high performance in clinics where access to experienced gastroenterologists is difficult. An accurate detection system can also improve the detection rate of smaller pre-cancerous polyps using the Convolutional Neural Network (CNN) models. The sensitivity of current colonoscopy suffers as the size of the colon becomes smaller. This can be improved because the state-of-the-art CNN models can extract features from objects at different scales.

5.1.1 Datasets Selection and Annotation

We have collected all publicly available endoscopic datasets in the research community, as well as collected a new dataset from the University of Kansas Medical Center. All datasets are deidentified without revealing the patient information. With the help of three endoscopists, we annotated the polyp classes of all collected video sequences and the bounding boxes of the polyp in every frame. Below is an introduction to each dataset.

MICCAI 2017 ***[8]***. This dataset is designed for Gastrointestinal Image Analysis (GIANA). It contains 18 videos for training and 20 videos for testing. The dataset is only labeled with polyp masks to test the ability to identify and localize polyps within images. There are no classification labels in this dataset. We converted the polyp masks into bounding boxes for each frame and annotated the polyp class.

CVC colon DB ***[9]***. The dataset has 15 short colonoscopy videos with a total of 300 frames. The labels are in the form of segmentation masks, and there are no classification labels. We extracted the bounding boxes and labeled the polyp class.

GLRC dataset ***[10]***. The Gastrointestinal Lesions in Regular Colonoscopy Dataset (GLRC) contains 76 short video sequences with class labels. There is no label for polyp location. We manually annotated the bounding box of each polyp frame by frame.

KUMC dataset. The dataset was collected from the University of Kansas Medical Center. It contains 80 colonoscopy video sequences. We manually labeled the bounding boxes as well as the polyp classes for the entire dataset

5.1.2 Dataset Build

The performance of a CNN model is highly dependent on the dataset. During training, a CNN model learns from many examples how to extract semantic features, on which localization and classification are based. Therefore, CNN detectors perform better when the dataset consists of representative examples of all categories. For example, images that are taken from different viewpoints, various illumination conditions, multiple sizes, etc. The more representative the dataset is, the more likely the CNN models can learn meaningful features for detection and classification. Then, at the inference time, the trained CNN models will have a higher ability to generalize the feature extraction on new input images.

In the research community, there are several small collections of endoscopic video datasets for different research purposes, such as MICCAI 2017, Gastrointestinal Lesions in Regular Colonoscopy Data Set (GLRC) ***[10]*** and CVC colon DB ***[9]*** dataset. However, after careful observation and analysis, we found that these datasets differ greatly from each other in terms of resolution and color temperature, as shown in **Fig.12**. This is largely due to the setups and characteristics of different imaging equipment used for data collection.

As pointed out, two of the main reasons why current CNN models perform worse in the real-world compared to benchmark test sets are the variance in image backgrounds and image quality. As shown in **Fig.12**, the images in different datasets vary greatly. If we train the models using only one of these datasets, the models may have poor generalization ability, and their performance will suffer when being applied to colonoscopy images from different devices in another medical facility.

(a) MICCAI (b) GLRC (c) KUMC

Fig.12 Samples frames from different colonoscopy datasets.
(a) has a higher resolution and a warm color temperature.
(b) has lower resolution and a green tone.
(c) is more natural in color tone but has a transparent cover around the frame edges.

To make the best use of the recent development of deep learning technologies for object detection and classification. We collected and created an endoscopic dataset and compared the performance of the state-of-the-art detectors for polyp detection and classification. These datasets come from various sources and serve different purposes as will be discussed in the earlier subsection. To integrate them together, we refer to PASCAL VOC object detection task to standardize the annotation. The dataset only contains two categories of polyps: **hyperplastic** and **adenomatous polyps**. It is important to train a model that could reliably differentiate them since adenomatous polyps are commonly considered as precancerous lesions that require resection while hyperplastic polyps are not.

5.1.3 Frame Selection

The video sequences from these datasets consist of different numbers of frames. For example, CVC colon DB only has 300 frames in total, averaging 20 frames per video sequence, while the number of frames in MICCAI 2017 varies from 400 to more than 1000 with a median value of around 300 in each sequence. The extreme imbalance among different lesions will reduce the

representativeness of the dataset. In addition, many frames in a long sequence are redundant since they are taken with very small camera movement. To avoid some long videos overwhelming others, we adopt an adaptive sampling rate to extract the frames from each video sequence based on the camera movement and video lengths to reduce the redundancy and homogenize the representativeness of each polyp. After sampling, we extracted around 300 to 500 frames for long sequences to maintain a balance among different sequences, while for small sequences like CVC colon DB, we simply keep all image frames in the sequence.

After extracting all frames, we carefully checked the generated dataset and manually removed some frames that contain misleading or not useful information. For example, when there is a sharp movement of the camera, the captured images may be severely blurred, out of focus, or subject to significant illumination change, as shown in **Fig.13**. These images cannot be accurately labeled, so they are removed. While some less flawed frames are kept improving the model's robustness under imperfect and noisy conditions.

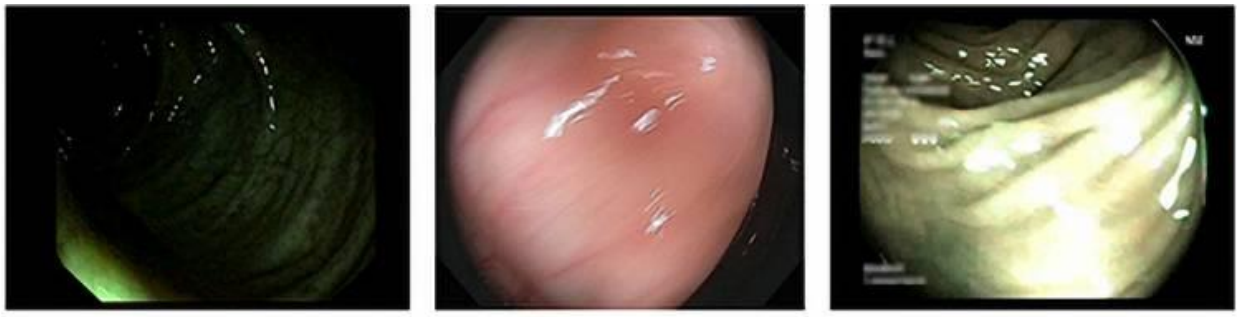


Fig.13 Some bad examples of colonoscopy frames.

Polyp classification only by visual examination is a big challenge, as reported in ***[10]***, the accuracy is normally below **70%** even for experienced endoscopists. In clinical practice, the results must be confirmed by further biopsy tests.

However, since we only have video sequences, when the endoscopist could not reach an agreement on the classification results, we simply remove those sequences from the dataset, otherwise, the models may not learn the correct information for classification. Eventually, the dataset contains 155 video sequences (37,899 image frames) with the labeled ground truth of the polyp classes and bounding boxes.

5.1.4 Dataset Split

To train and evaluate the performance of different learning models, we need to divide the combined dataset into training, validation, and test sets. For most benchmark datasets for generic object detection, the split is normally based on images. However, this does not apply to the endoscopic dataset. Because all frames in one video sequence correspond to the same polyp, if we split the dataset at the image level, then the same polyp will simultaneously appear in the training, validation, and test sets. This will falsely increase the classification performance since the models have already seen the polyps to be tested during the training stage. Therefore, we split the dataset at the video level.

Since the final dataset is combined from four different datasets captured by different equipment with different data distribution. To increase the representativeness of the dataset, as well as the balance of the two classes of the polyps, we make the division for each dataset and polyp class independently. For each class in one dataset, we randomly select 75%, 10%, and 15% sequences to form the training, validation, and test sets, respectively. For example, the GLRC ***[10]*** has 41 videos, with 26 adenomatous and 15 hyperplastic sequences. We split the 26 adenomatous sequences and the 15

hyperplastic sequences independently according to the same ratio to guarantee the class balance in the final dataset.

In summary, we have generated 116 training, 17 validation, and 22 test sequences, with 28773, 4254, and 4872 frames, respectively, for each set. Some sample frames from the dataset are shown in **Fig.14**. For the training set, we combine all frames from the 116 sequences into one folder and shuffle them.

While for the validation and test sets, we keep the sequence split to evaluate the model performance based on polyps (i.e., sequences). The details of the dataset organization are shown in **Table 1**. The dataset can be accessed from here ***[11]***.

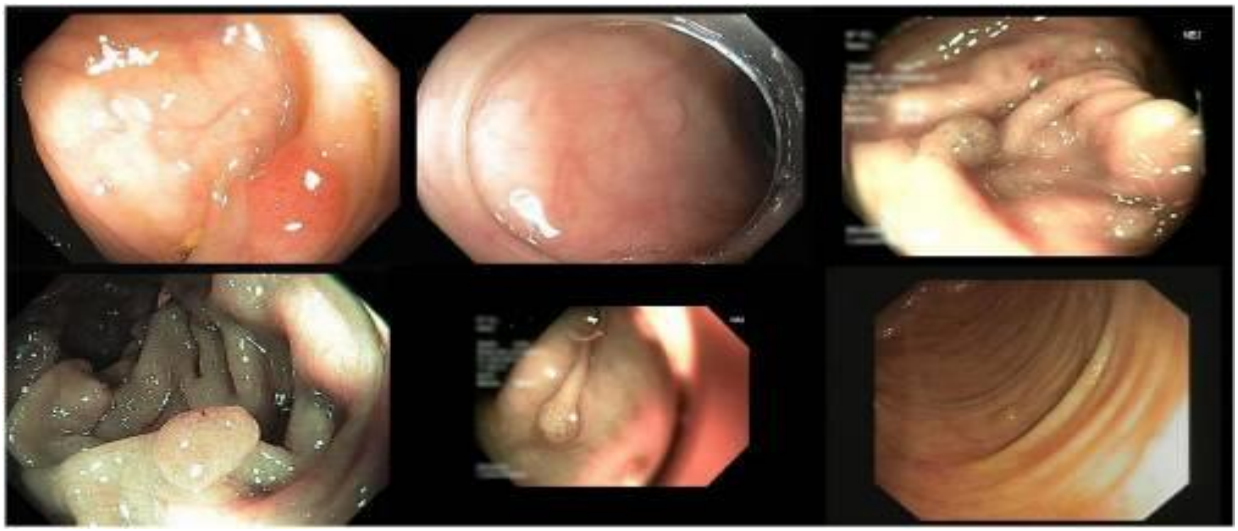


Fig.14 Six example frames from the generated dataset.

Table 1. Dataset Frame Selection and Split:

Dataset	MICCAI 2017	CVC Colon DB	GLRC	KUMC
Before selection	38	15	76	80
After selection	23	15	41	76
# of Seqs	**ad** **hp** 13 10 train val test 20 1 2 ad hp ad hp ad hp 12 8 0 1 1 1	**ad** **hp** 10 5 train val test 11 1 3 ad hp ad hp ad hp 7 4 1 0 2 1	**ad** **hp** 26 15 train val test 29 5 7 ad hp ad hp ad hp 19 10 3 2 4 3	**ad** **hp** 38 38 train val test 56 10 10 ad hp ad hp ad hp 27 29 5 5 6 4

➔ **ad**: Adenomatous **hp**: Hyperplastic

5.1.5 Further Data Preprocessing

Firstly, we carefully checked that every annotation in the Pascal VOC format correctly corresponds to the image label by going through all the images and annotations and matching them with each other. While checking for the correctness of the data we noticed some images that didn't have a corresponding label and annotation, so we removed all the images and annotation files that didn't have any labels from training, validation, and testing datasets.

Secondly, since the state-of-the-art models contain YOLO ***[12]*** models, we needed to convert the Pascal VOC format annotations to YOLO format annotation for all the images. After converting to YOLO format, we checked all the annotated images to validate that the bounding boxes were correct and accurate.

Lastly, we created train.txt, val.txt, and test.txt for YOLO V3 and YOLO V4 models. Also, we created train.tfrecords, val.tfrecord, and test.tfrecords for the state-of-the-art models following the steps in here:https://tensorflow-object-detection-api-tutorial.readthedocs.io/en/latest/install.html for installation and setting up the environment and the models.

In summary, we end up with **27047, 4213, 4719** frames for training, validation, and testing, respectively.

5.1.6 Evaluation Models for Detection and/or Classification

5.1.6.1 Classification Models:

Mobile-net V2 ***[13]*****:** is a convolutional neural network architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure

where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depth wise convolutions to filter features as a source of non-linearity. The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers.

Inception V3, V2, V1 ***[14 – 17 – 18]*****:** Inception Modules are incorporated into convolutional neural networks (CNNs) as a way of reducing computational expense. As a neural net deals with a vast array of images, with wide variation in the featured image content, also known as the salient parts, they need to be designed appropriately. The most simplified version of an inception module works by performing a convolution on an input with not one, but three different sizes of filters (1x1, 3x3, 5x5). Also, max pooling is performed. Then, the resulting outputs are concatenated and sent to the next layer. By structuring the CNN to perform its convolutions on the same level, the network gets progressively wider, not deeper.

To make the process even less computationally expensive, the neural network can be designed to add an extra 1x1 convolution before the 3x3 and 5x5 layers. By doing so, the number of input channels is limited and 1x1 convolutions are far cheaper than 5x5 convolutions.

General-cifar10-covnet ***[15]*****:** Covnet is a general network architecture for deep learning that was trained on the cifar10 dataset.

CNNs are particularly useful for finding patterns in images to recognize objects, classes, and categories. They can also be quite effective for classifying audio, time-series, and signal data.

VGG16 and VGG19 ***[16]*****:** VGG stands for Visual Geometry Group; it is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The "deep" refers to the number of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers. The VGG architecture is the basis of ground-breaking object recognition models. Developed as a deep neural network, the VGGNet also surpasses baselines on many tasks and datasets beyond ImageNet. Moreover, it is now still one of the most popular image recognition architectures.

Inception ResNet V2 ***[19]*****:** Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database *[1]*. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 299-by-299.

Mobile-Net-Edge TPU V2 ***[20]*****:** has the same architecture as MobileNet V2 *[13]*. But MobileNetEdgeTPU achieves higher accuracy and lower latency compared with other models.

ResNet V1_50, V2_152 and V2_101 ***[21 – 22 – 23]*****:** ResNet uses Batch Normalization at its core. The Batch Normalization adjusts the input layer to increase the performance of the network. The problem of covariate shift is mitigated. ResNet makes use of the Identity Connection, which helps to protect the network from vanishing gradient problem.

Deep Residual Network uses bottleneck residual block design to increase the performance of the network. The ResNet V2 mainly focuses on making the second non-linearity as an identity mapping i.e., the output of addition operation

between the identity mapping and the residual mapping should be passed as it is to the next block for further processing. However, the output of the addition operation in ResNet V1 passes from ReLU activation and then transferred to the next block as the input. The models architectures also differ in the number of their layers 50, 101, 152 respectively.

5.1.6.2 Object Detection Models:

SSD ResNet50 V1 FPN 640*640 (RetinaNet50) ***[24]*****:** Single Shot Detector (SSD), as one of the most successful one-stage detectors, has become the foundation of many other studies. It takes advantage of the different sizes of feature maps and utilizes a simple architecture to generate predictions at different feature map scales. SSD can achieve a fast detection rate with competitive accuracy.

Faster R-CNN ResNet50 V1 640*640 ***[24]*****:** Faster RCNN is a two-stage framework model and one of the families of RCNN networks. It improves the Fast RCNN network by replacing the slow selective search algorithm with a region proposal network, resulting in a faster detection rate. Furthermore, the region proposal network is trainable, which can potentially achieve better performance.

YOLO V3 ***[25]*****:** YOLOv3 is an iterative improvement of YOLO (You Only Look Once). It improves the performance of its previous versions by introducing a new backbone network, multi-scale prediction, and a modified class prediction loss function.

YOLO V4 ***[25]*****:** YOLOv4 explores the bag of freebies and bag of specials and selects some of them in the new detection model. The basic rules for a detection

model are high-resolution input images for detecting relatively small objects, deeper layers for a larger receptive field, and more parameters for detecting various objects. Based on those rules, YOLOv4 selects various effective bag of freebies and bag of specials to enhance the performance of the model while maintains high-speed inference. In addition, instead of exploiting DarkNet53 as the backbone in YOLOv3, an enhanced version of DarkNet53 (CSPDarknet53) is selected as the backbone for YOLOv4.

YOLO V5: YOLOv5 is a model in the You Only Look Once (YOLO) family of computer vision models. YOLOv5 is commonly used for detecting objects. YOLOv5 comes in four main versions: small (s), medium (m), large (l), and extra-large (x), each offering progressively higher accuracy rates. Each variant also takes a different amount of time to train.

YOLO V8 ***[26]*****:** is the latest version of the YOLO object detection and image segmentation model developed by Ultralytics. YOLOv8 is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility**.**

One key feature of YOLOv8 is its extensibility. It is designed as a framework that supports all previous versions of YOLO, making it easy to switch between different versions and compare their performance.

In addition to its extensibility, YOLOv8 includes a few other innovations that make it an appealing choice for a wide range of object detection and image segmentation tasks. These include a new backbone network, a new anchor-free detection head, and a new loss function. YOLOv8 is also highly efficient and can be run on a variety of hardware platforms, from CPUs to GPUs.

5.1.7 Experiments

5.1.7.1 Frame-based two-class Polyps Detection

This experiment predicts polyps for individual frames. It is a test of a model's localization and classification ability. The CNN models are trained using our training set that consists of a mix of frames from different video sequences. During the validation and test phase, we treat each frame individually and evaluate the performance.

To set the benchmark performance we trained the models on Google Colab with Tesla T4 16 GB GPU and 16 GB ram.

5.1.7.2 Polyps Classification

With the generated dataset, we evaluated the earlier mentioned classification models, including MobileNet, Inception, Covnet, VGG16, VGG19, Inception-ResNet, MobileNet-Edge, and ResNet.

To set the benchmark performance we trained the models on HP G3 intel i7 core with NVIDIA GeForce GTX 1060 6 GB and 16 GB RAM laptop for 10 epochs with input size of 224*224 or 299*299 depending on the model architecture with Adam optimizer and SparseCategoricalCrossentropy as a loss function setting "from_logits" parameter to True and a final dense layer of 2 neurons for classification.

5.1.7.3 Frame-based and Sequence-based one-class Polyps Detection

This experiment has almost the same setups as the frame-based two-class polyp detection except for the class number. Hyperplastic and adenomatous polyps are treated as single class *Polyps*.

In colorectal cancer screening, it is more important to accurately detect whether polyps are developed than classifying polyp categories, because further screening and diagnosis are always followed after colonoscopy finds suspected lesions. This experiment aims to test whether a higher performance could be achieved by only localizing polyps in general. Without the more challenging task of classifying polyp categories, which has been already conducted in our experiments.

To set the benchmark performance we trained the models on Google Colab with Tesla T4 16 GB GPU and 16 GB ram.

5.1.7.4 Classification of Endoscopic and Random Images

In this section we decided to build a classifier to differentiate between endoscopic images and false entry images for the purpose of verifying and validating that the entered image fall into the endoscopic images category for our object detection model to work as it supposed to.

We trained the model on HP G3 intel i7 core with NVIDIA GeForce GTX 1060 6 GB and 16 GB RAM laptop.

We used the entire Polyps dataset and a random dataset that is collected from 8 random datasets containing images for cars, birds, landscape, people, traffic, forests, colors, and food.

We used Inception V3 for training our model with image size of 299*299 and batch size of 16 and Adam optimizer with learning rate of 0.0001.

5.1.8 Results and Analysis

5.1.8.1 The results for frame-based two-class Polyps detection

The results for training for 40k iterations (40 epochs) and testing on the test set are shown in **Table 2**.

Table 2. Frame-Based Two-Class Polyps Training Results:

Model:	SSD ResNet50	Faster R-CNN ResNet50	YOLO V3	YOLO V4	YOLO V5	YOLO V8
mAP:	**50.2%**	**52.7%**	**48.4%**	**54.6%**	**55.03%**	**54.9%**

➔ **mAP**: **Mean Average Precision**

With the low mAP, we need to try a different approach for training.

5.1.8.2 The results for the Polyps classification experiment

Table 3. Polyps Classification Results:

Model	Train-accuracy	Train-loss	Val-accuracy	Val-loss
MobileNet V2	47.52 %	0.6931	39.47 %	0.6931
Inception V3	94.14 %	0.1677	55.62 %	1.1491
Inception V2	91.26 %	0.2334	57.97 %	0.9461
Inception V1	91.54 %	0.2319	60.51 %	0.9076
Covnet	43.55%	0.7564	40.33 %	1.2222
VGG16	97.64 %	0.0125	43.79 %	2.9950
VGG19	**99.80 %**	**0.0085**	42.29 %	3.1780
Inception-ResNet V2	97.13 %	0.0996	**67.20 %**	0.8189
MobileNet-Edge	51.71 %	0.0857	50.28 %	1.3701
ResNet V1-50	97.08 %	0.0893	58.77 %	1.2333
ResNet V2-101	95.85 %	0.1183	53.26 %	1.6859
ResNet V2-152	94.78	0.1201	51.23%	1.7962

As seen from the results above the models have poor classifying properties and only the training accuracy is high, so the models are memorizing the dataset instead of generalizing (Overfitting). We concluded that the models are too complex for the dataset to be classified, so we decided to make our own model with **Inception V4 backbone**.

5.1.8.2.1 The Proposed Model

We took some of **Inception V4** architecture and reduced the image size to 28*28 and built our neural network model with the following model summary. **Fig.15**:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 28, 28, 20)        1520

 activation (Activation)     (None, 28, 28, 20)        0

 max_pooling2d (MaxPooling2D  (None, 14, 14, 20)       0
 )

 conv2d_1 (Conv2D)           (None, 14, 14, 50)        25050

 activation_1 (Activation)   (None, 14, 14, 50)        0

 max_pooling2d_1 (MaxPooling  (None, 7, 7, 50)         0
 2D)

 flatten (Flatten)           (None, 2450)              0

 dense (Dense)               (None, 500)               1225500

 activation_2 (Activation)   (None, 500)               0

 dense_1 (Dense)             (None, 2)                 1002
...
Total params: 1,253,072
Trainable params: 1,253,072
Non-trainable params: 0
_________________________________________________________________
```

Fig.15 The Proposed Model Summary.

The proposed model achieved a training and validation accuracy of **98.46%** and **99.07%**, respectively after training with 40 epochs and with batch size of 16 and an initial learning rate of 0.001. Also, it achieved **97.69%** accuracy on the test set.

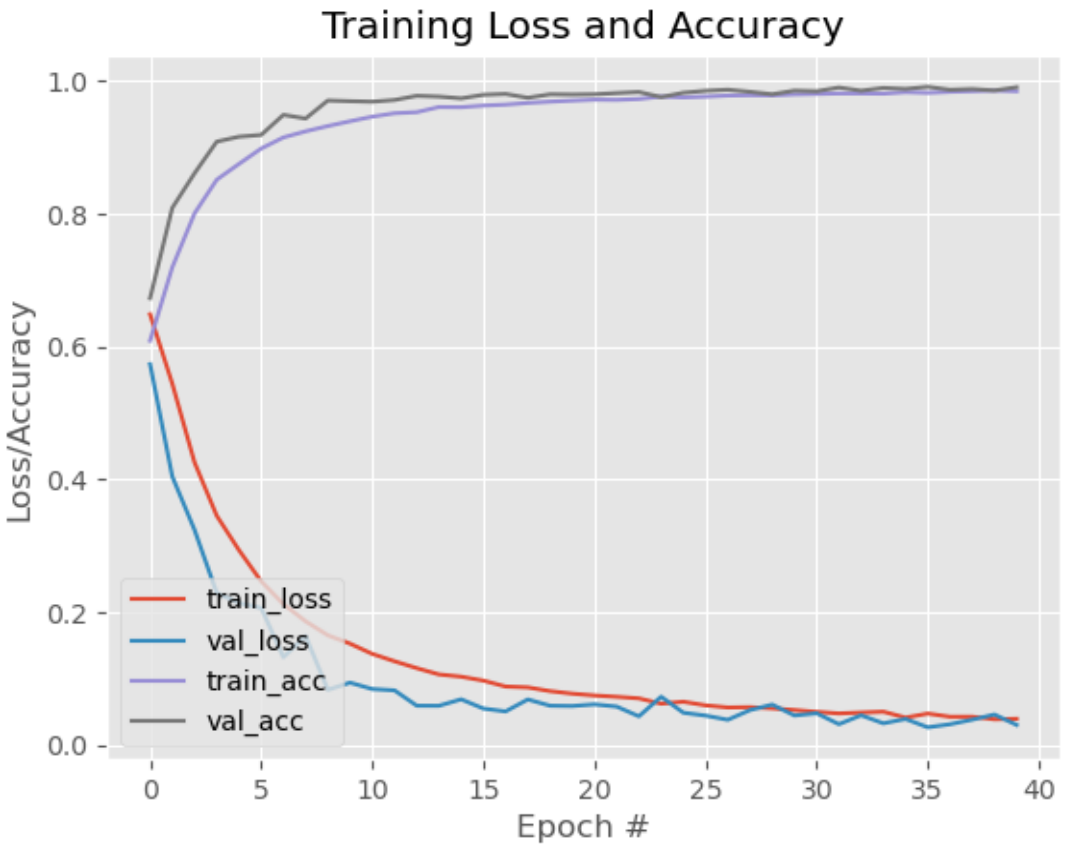


Fig.16 The Model's Training Graph.

5.1.8.3 The results for frame-based one-class Polyp detection testing

After experimenting different models and model sizes for 50 epochs and testing them on our test set, here are the best models results shown in **Table 4**:

Table 4. Sequence-Based One-Class Polyp Testing Results:

Class	Model	Image size	Batch size	Box Precision	Recall	mAP
Adenomatous	**Yolo v8 M**	**416**	**32**	**84.3%**	**68.4%**	**75.8%**
Hyperplastic	**Yolo v8 M**	**416**	**24**	**83.8%**	**79.3%**	**79.7%**

5.1.8.3 The results for sequence-based one-class Polyp detection testing

We tested on the sequences of frames in each folder, we concluded that the images with light background were detected with much higher precision than images with darker background.

5.1.8.3.1 Adenomatous Model

Some confusion matrices of light backgrounded images and their examples are presented below in **Fig.17** and **Fig.18**:

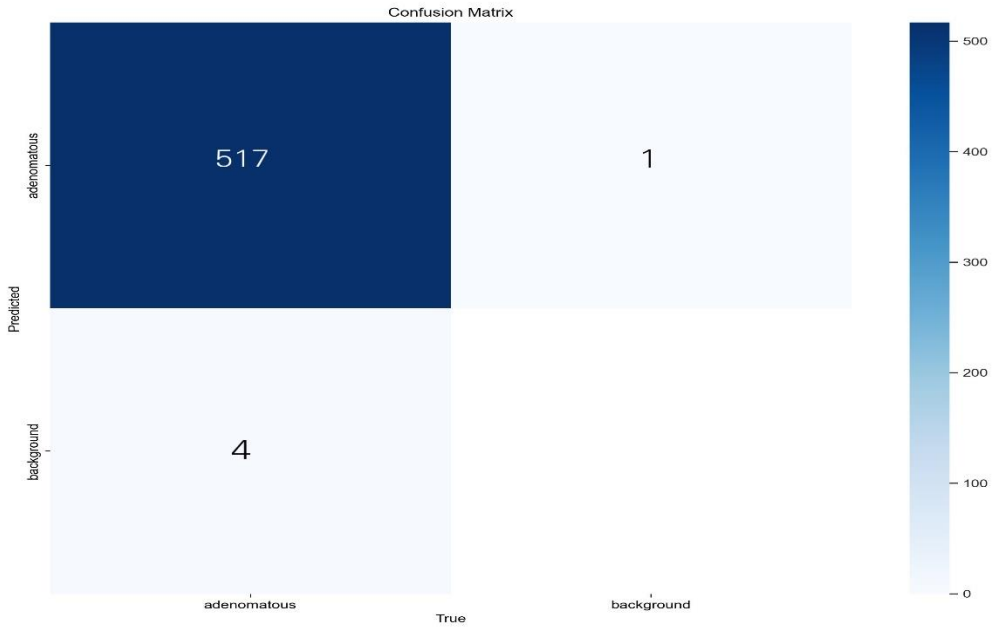
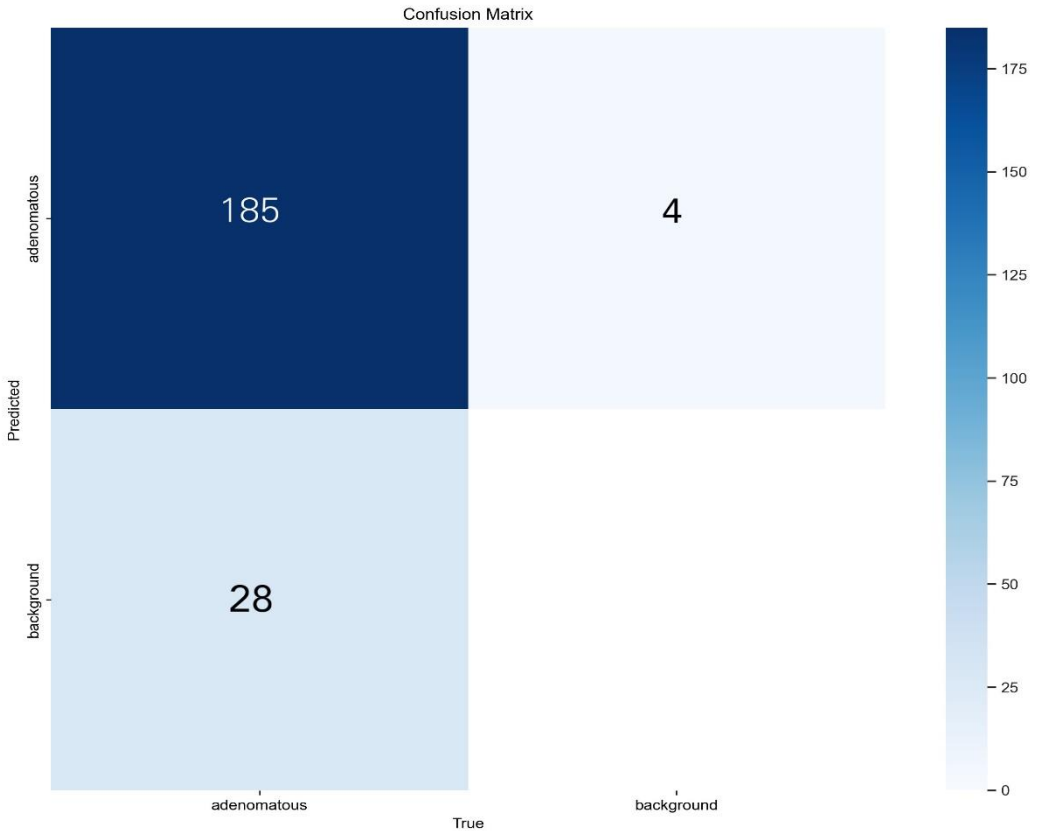


Fig.17 Confusion Matrix and Examples of Light Backgrounded Images.



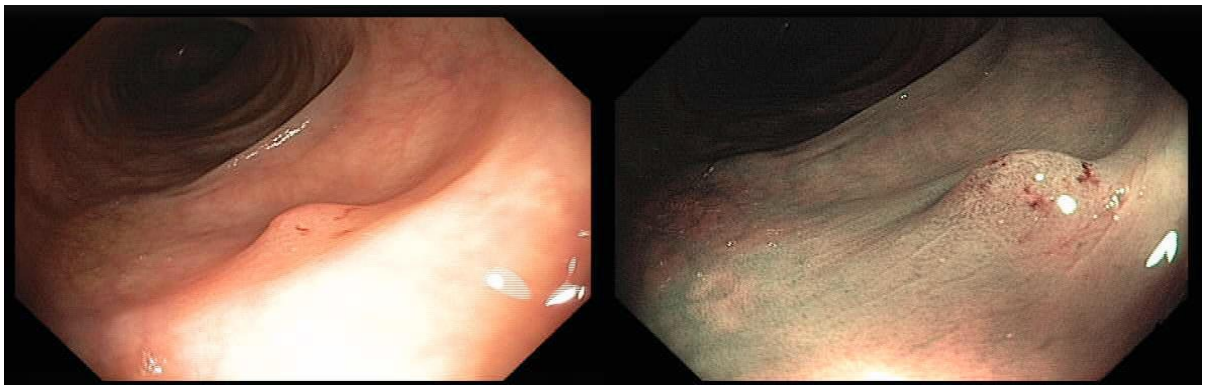


Fig.18 Confusion Matrix and Examples of Light Backgrounded Images.

Confusion matrix of dark backgrounded images and their examples are presented below in **Fig.19**:

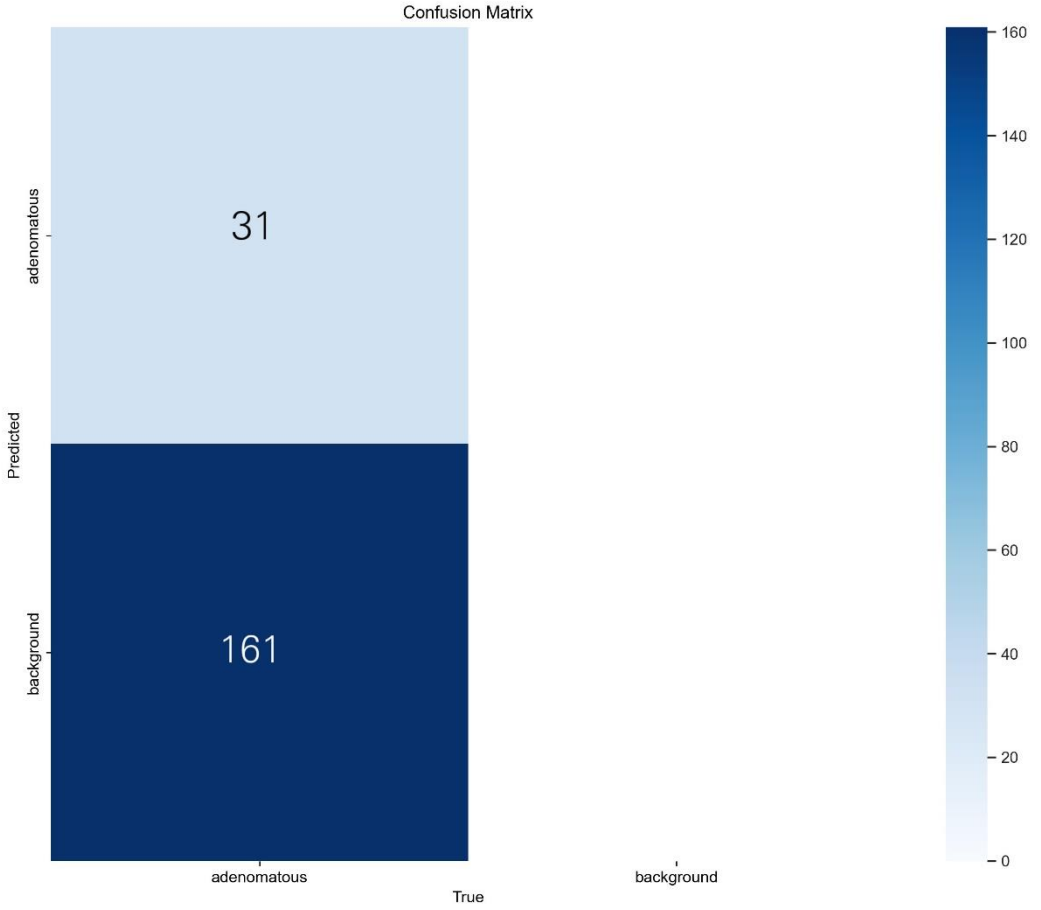
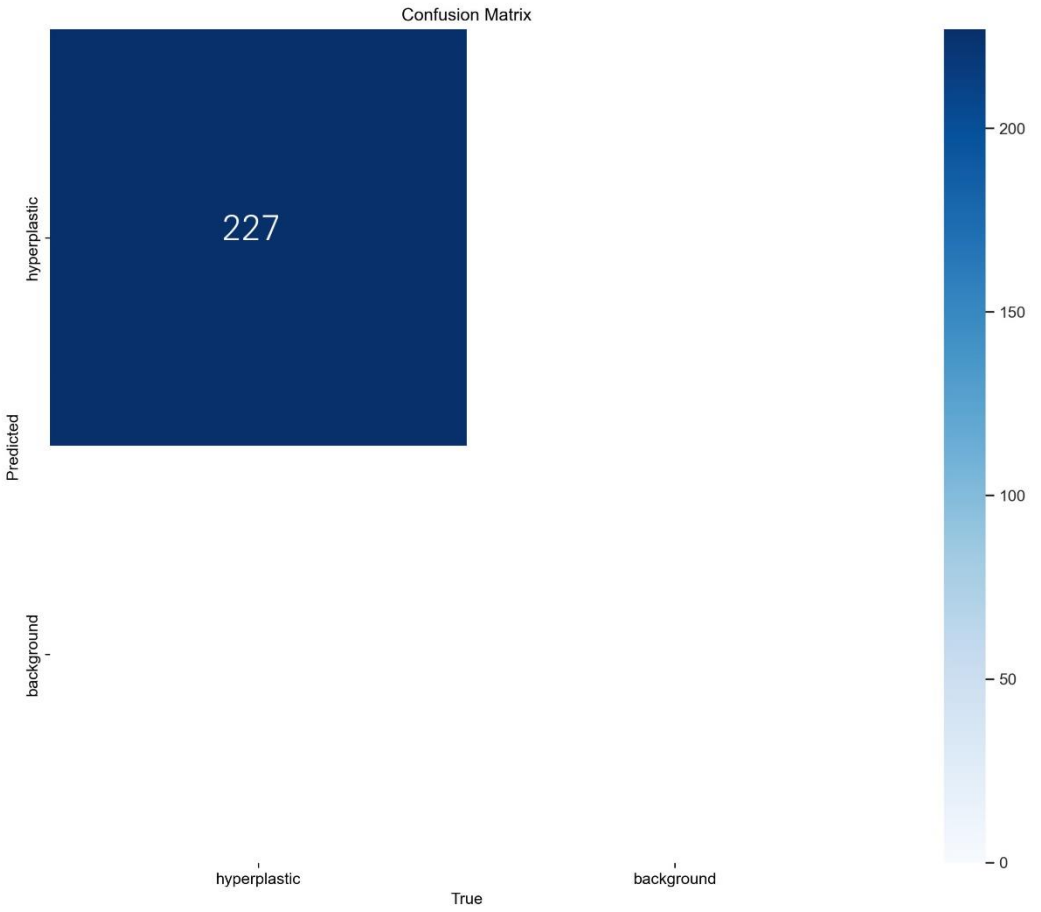


Fig.19 Confusion Matrix and Examples of Dark Backgrounded Images.

5.1.8.3.2 Hyperplastic Model

Confusion matrix of light backgrounded images and their examples are presented below in **Fig.20**:



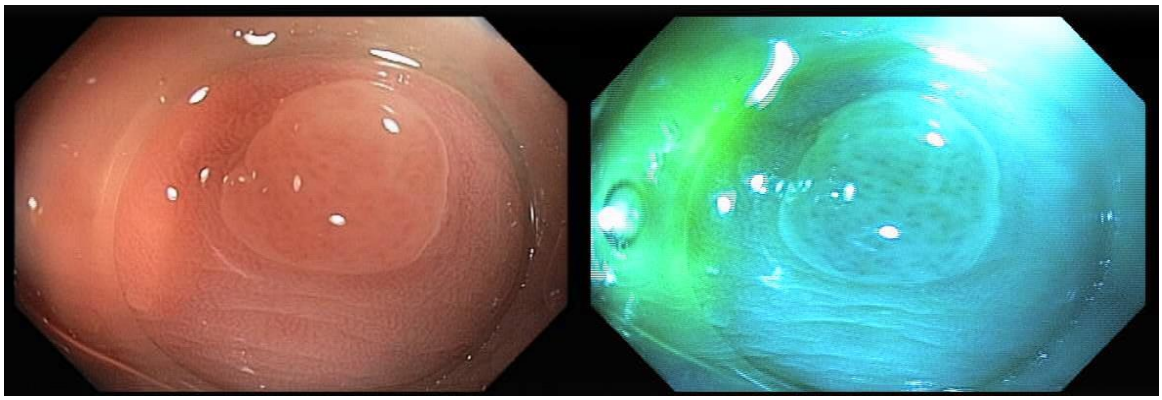
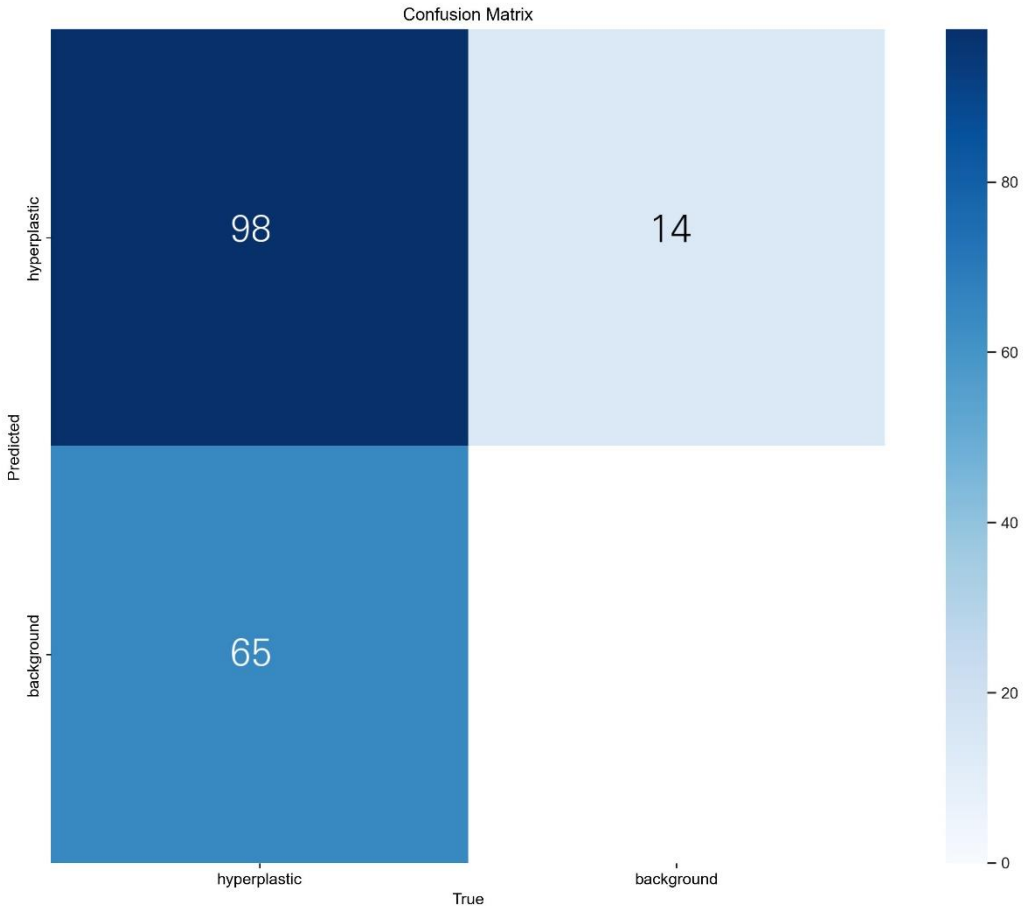


Fig.20 Confusion Matrix and Examples of Light Backgrounded Images.

Confusion matrix of dark backgrounded images and their examples are presented below in **Fig.21**:



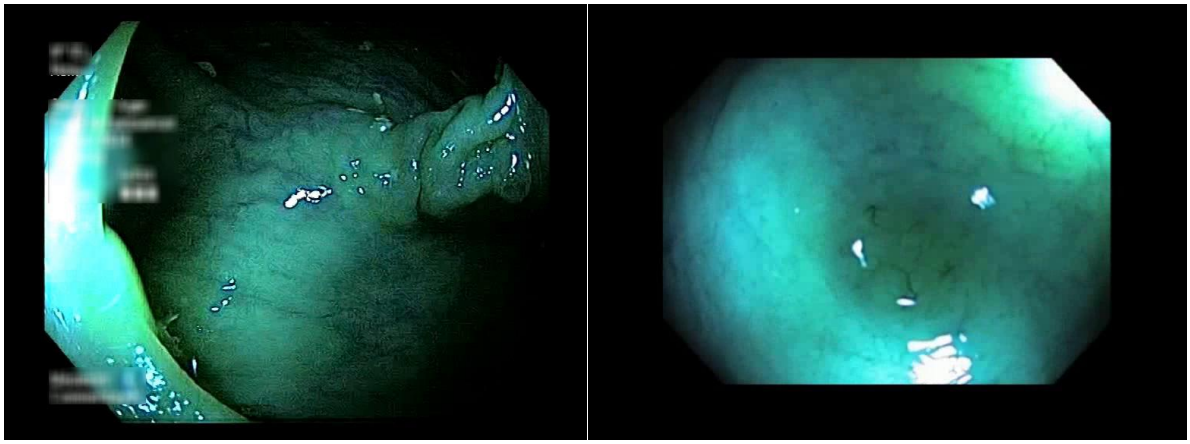


Fig.21 Confusion Matrix and Examples of Dark Backgrounded Images.

5.1.8.4 The results for classification of endoscopic and random images

After training our model for 10 epochs, we have achieved **100%** accuracy on the training set, and **99.99%** accuracy on the validation set, and **99.99%** accuracy on the test set.

The model's training graph in **Fig.22**:

Fig.22 Inception V3 Training Graph.

5.1.8.4.1 The approach for training our classifier:

Firstly, we pass the image into our Endoscopic Classification model to ensure that the entered image is an endoscopic image, otherwise it won't continue to our next model. This improves our classifier accuracy and prevent false data entry and reduces the inference time.

Secondly, we pass the endoscopic image to our Polyps classification model that classifies in as Benign (Hyperplastic) or Malignant (Adenomatous) and returns the classification confidence, then if the image is Benign it is redirected to our Hyperplastic polyp detection model after asking the user if he wants to proceed to detecting the polyp in the image, and if the image is Malignant it is redirected to our Adenomatous polyp detection model. If the user chooses to not proceed with the polyp detection, then the result of the Polyps classification model will be the output on the screen and our classifier terminates, that reduces the time complexity and inference time.

Below in **Fig.23**, the flowchart of the classifier sequence:

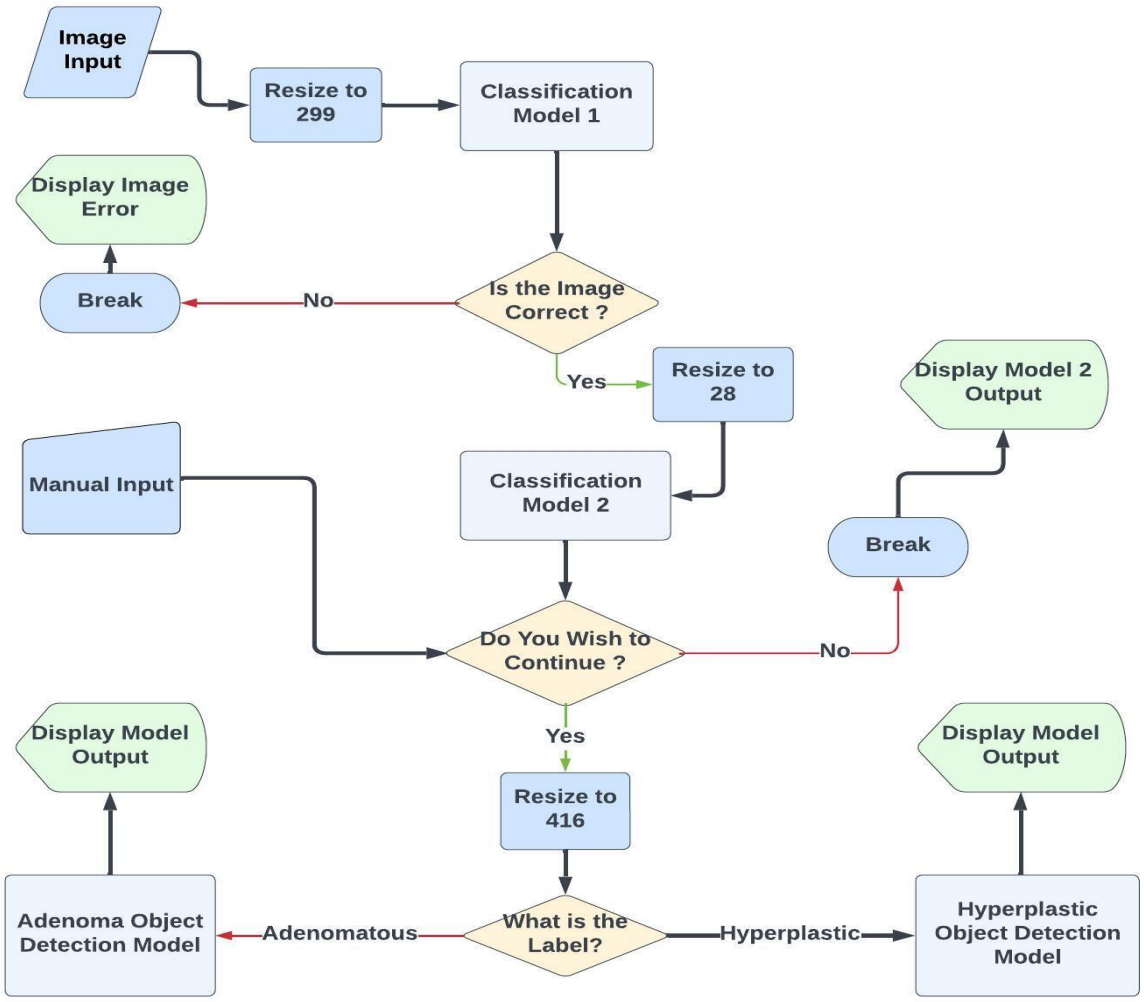


Fig.23 The Flowchart of the Endoscopy Classifier.

5.2 Classification of Histopathological adenocarcinoma and benign epithelium

Classification of colon cancer has great importance in medical diagnosis. An accurate and real-time investigation provides medical experts to review typical treatment timely. Histopathological image analysis is the gold standard for pathologists to grade colorectal cancers of different differentiation types. In recent years, with the increasing number of colorectal cancer patients, the workload of physicians is increasing day by day. In addition, the low differentiation of histopathological images of colorectal cancers of different differentiation types makes the diagnosis complicated and time-consuming, which may lead to misdiagnosis and missed diagnosis. Although gastroenterology clinics have a high demand for colon specimens, pathologists have a long training period (> 10 years) ***[27]***. According to the Chinese Association of Pathologists, China, a country of 1.4 billion people, has only 20,000 professionally accredited pathologists ***[28]***. Therefore, it is particularly important to build an efficient computerized automatic diagnostic model to effectively identify histopathological images of colorectal cancer at multiple levels, and then assist pathologists in objective diagnosis and grading.

The recent World Health Organization (WHO) classification for malignant epithelial tumors of the colorectum includes four main categories: adenocarcinoma (ADC) not otherwise specified (NOS), neuroendocrine tumor NOS, neuroendocrine carcinoma NOS and mixed neuroendocrine-non-neuroendocrine neoplasm (MiMEN) ***[29]***. Of these, colorectal ADC is the most common (90%) and, by definition, it shows glandular and mucinous differentiation. The diagnosis of CRC is only the first step for a complete pathology report.

Histopathology image generations start with the standard procedure of tissue preparation. Biopsy or surgical specimens (representative sections) are formalin-fixed and paraffin-embedded. Then, the 4µm tissue sections are prepared and stained with H&E dye. The images are extracted after a scanning procedure.

5.2.1 Datasets Selection

We have collected three of the publicly available datasets. Below is an introduction to each dataset.

NCT-CRC-HE-100K *[30]*: This is a set of 100,000 non-overlapping image patches from hematoxylin & eosin (H&E) stained histological images of human colorectal cancer (CRC) and normal tissue.

All images are 224x224 pixels (px) at 0.5 microns per pixel (MPP). All images are color-normalized using Macenko's method (http://ieeexplore.ieee.org/abstract/document/5193250/.DOI 10.1109/ISBI.2009.5193250).

Tissue classes are: Adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), colorectal adenocarcinoma epithelium (TUM).

These images were manually extracted from N=86 H&E-stained human cancer tissue slides from formalin-fixed paraffin-embedded (FFPE) samples from the NCT Biobank (National Center for Tumor Diseases, Heidelberg, Germany) and the UMM pathology archive (University Medical Center Mannheim, Mannheim, Germany). Tissue samples contained CRC primary tumor slides and

tumor tissue from CRC liver metastases; normal tissue classes were augmented with non-tumorous regions from gastrectomy specimen to increase variability.

CRC-VAL-HE-7K *[30]*: This is a set of 7180 image patches from N=50 patients with colorectal adenocarcinoma (no overlap with patients in NCT-CRC-HE-100K). It can be used as a validation set for models trained on the larger data set. Like in the larger data set, images are 224x224 (px) at 0.5 MPP. All tissue samples were provided by the NCT tissue bank, see above for further details and ethics statement.

LC25000 *[31]*: The dataset contains color 25,000 images with 5 classes of 5,000 images each. All images are 768 x 768 pixels in size and are in jpeg file format. Our dataset can be downloaded as a 1.85 GB zip file LC25000.zip. After unzipping, the main folder lung-colon-image-set contains two subfolders: colon-image-sets and lung-image-sets.

The subfolder colon-image-sets contains two secondary subfolders: colon-aca subfolder with 5000 images of colon adenocarcinomas and colon-n subfolder with 5000 images of benign colonic tissues.

The subfolder lung-image-sets contains three secondary subfolders: lung-aca subfolder with 5000 images of lung adenocarcinomas, lung-scc subfolder with 5000 images of lung squamous cell carcinomas, and lung-n subfolder with 5000 images of benign lung tissues.

Random images dataset: we collected 8 random datasets containing images for cars, birds, landscape, people, traffic, forests, colors, and food for false entry.

5.2.2 Datasets Build

Since our interest is to build a model for classifying colorectal adenocarcinoma epithelium tissues and normal colon tissues, we only needed two subfolders

from **NCT-CRC-HE-100K** ***[30]*** **and CRC-VAL-HE-7K** ***[30]*** **which are** colorectal adenocarcinoma epithelium **(TUM) and** normal colon mucosa **(NORM), and from LC25000** ***[31]*** dataset we only needed the subfolder colon-image-sets that contains colon adenocarcinoma and benign colonic tissues.

We will use the **NCT-CRC-HE-100K** ***[30]*** **dataset and LC25000** ***[31]*** dataset for training and validation only, while the **CRC-VAL-HE-7K** ***[30]*** dataset will be used for testing for better generalization.

5.2.3 Datasets Split

Originally, **NCT-CRC-HE-100K** ***[30]*** dataset has 14317, 8763 images for colorectal adenocarcinoma (TUM) and normal colonic tissues (NORM), respectively. While the **LC25000** ***[31]*** dataset have 5000, 5000 images for colorectal adenocarcinoma (colon-aca) and normal colonic tissues (colon-n), respectively.

We combined both datasets, so it resulted in 19317, 13763 images for colorectal adenocarcinoma, and normal colon classes, respectively. Which resulted in 33080 images for both classes. Then we split the generated dataset into train, validation by the ration (90, 10).

The **CRC-VAL-HE-7K** ***[30]*** dataset has 1233, 741 images for colorectal adenocarcinoma (TUM) and normal colonic tissues (NORM), respectively. And it will be used for testing.

In summary, we have **17386** adenocarcinoma, **12387** normal images for training, and **1931** adenocarcinoma, **1376** normal images for validation, and the test set.

In the **Random images dataset,** we collected 26376 images of random images and split them into train, validation, and test by the ratio (70, 20, 10), which resulted in **18463**, **5275**, **2638** images for train, validation, and test, respectively.

5.2.4 Evaluation Models for Classification

Inception ResNet V2 ***[19]*****:** Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database *[1]*. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 299-by-299.

Inception V3, V2, V1 ***[14 – 17 – 18]*****:** Inception Modules are incorporated into convolutional neural networks (CNNs) as a way of reducing computational expense. As a neural net deals with a vast array of images, with wide variation in the featured image content, also known as the salient parts, they need to be designed appropriately. The most simplified version of an inception module works by performing a convolution on an input with not one, but three different sizes of filters (1x1, 3x3, 5x5). Also, max pooling is performed. Then, the resulting outputs are concatenated and sent to the next layer. By structuring the CNN to perform its convolutions on the same level, the network gets progressively wider, not deeper.

EfficientNetV2 ***[32]*****:** is a type of convolutional neural network that has faster training speed and better parameter efficiency than previous models. To develop these models, the authors use a combination of training-aware neural architecture search and scaling, to jointly optimize training speed. The models

were searched from the search space enriched with new ops such as Fused-MBConv.

Architecturally the main differences are:

- EfficientNetV2 extensively uses both MBConv and the newly added fused-MBConv in the early layers.
- EfficientNetV2 prefers smaller expansion ratio for MBConv since smaller expansion ratios tend to have less memory access overhead.
- EfficientNetV2 prefers smaller 3x3 kernel sizes, but it adds more layers to compensate the reduced receptive field resulted from the smaller kernel size.
- EfficientNetV2 completely removes the last stride-1 stage in the original EfficientNet, perhaps due to its large parameter size and memory access overhead.

ResNet V1_101 and V2_101 ***[22 – 23]*****:** ResNet uses Batch Normalization at its core. The Batch Normalization adjusts the input layer to increase the performance of the network. The problem of covariate shift is mitigated. ResNet makes use of the Identity Connection, which helps to protect the network from vanishing gradient problem.

Deep Residual Network uses bottleneck residual block design to increase the performance of the network. The ResNet V2 mainly focuses on making the second non-linearity as an identity mapping i.e., the output of addition operation between the identity mapping and the residual mapping should be passed as it is to the next block for further processing. However, the output of the addition operation in ResNet V1 passes from ReLU activation and then transferred to the next block as the input.

5.2.5 Experiments

We decided to conduct two experiments in training the models:

- Classifying adenocarcinoma and benign tissues only (binary classification)

- Classifying adenocarcinoma, benign tissues, and false entry class (multi-class classification)

5.2.5.1 Classifying adenocarcinoma and benign tissues only (binary classification)

The generated dataset has **29773**, **3307**, and **1974** for train, validation, and test, respectively. With the generated dataset, we evaluated the classification models accuracy. Including Inception, VGG16, VGG19, and Inception-ResNet. To set the benchmark performance we trained the models on HP G3 intel i7 core with NVIDIA GeForce GTX 1060 6 GB and 16 GB RAM laptop for 21 epochs and batch size of 16 to 8 depending on the model size with input size of 224*224 or 299*299 depending on the model architecture with StochasticGradientDescent optimizer setting the “learning_rate” and “momentum” parameters to 0.001, 0.9, respectively. We used SparseCategoricalCrossentropy as a loss function setting “from_logits” parameter to True. We also added a dense layer with 512 neurons after the keras model layer for Inception-ResNet model only, and a final dense layer of 2 neurons for classification.

5.2.5.2 Classifying adenocarcinoma, benign tissues, and false entry class (multi-class classification)

The generated dataset has **48236**, **8582**, and **4612** for train, validation, and test, respectively. With the generated dataset, we evaluated the classification models

accuracy. Including Inception, and Inception-ResNet. To set the benchmark performance we trained the models on HP G3 intel i7 core with NVIDIA GeForce GTX 1060 6 GB and 16 GB RAM laptop for 20 epochs and batch size of 16 to 8 depending on the model size with input size of 224*224 or 299*299 depending on the model architecture with StochasticGradientDescent optimizer setting the "learning_rate" and "momentum" parameters to 0.001, 0.9, respectively. We used SparseCategoricalCrossentropy as a loss function setting "from_logits" parameter to True. We also added a dense layer with 512 neurons after the keras model layer for Inception-ResNet model only, and a final dense layer of 3 neurons for classification.

5.2.6 Results

5.2.6.1 Classifying adenocarcinoma and benign tissues only (binary classification)

The results of testing the models with batch size of 16 are shown in **Table 5**

Table 5. Binary Classification Testing Results:

Model	Loss	Accuracy	Inference time
Inception-ResNet V2	0.0994	96.81 %	39s
Inception V3	**0.0701**	**97.16 %**	17s
Inception V1	0.1251	95.49 %	8s
Inception V2	0.0951	96.81 %	9s
EfficientNet-V2-imagenet21k-s	0.1894	95.64 %	47s
ResNet V1_101	0.1235	96.30 %	19s
ResNet V2_101	0.1260	95.85 %	22s

5.2.6.2 Classifying adenocarcinoma, benign tissues, and false entry class (multi-class classification)

The results of testing the models with batch size of 16 are shown in **Table 6**

Table 6. Multi-Class Classification Testing Results:

Model	Loss	Accuracy	Inference time
Inception-ResNet V2	0.0457	98.55 %	76s
Inception V3	**0.0375**	**98.76 %**	40s
Inception V1	0.0577	97.79 %	20s
Inception V2	0.0486	98.37 %	**17s**
EfficientNet-V2-imagenet21k-s	0.0769	98.32	101s
ResNet V1_101	0.0501	98.31 %	38s
ResNet V2_101	0.0606	98.20 %	49s

The model's training graph presented in **Fig.24:**

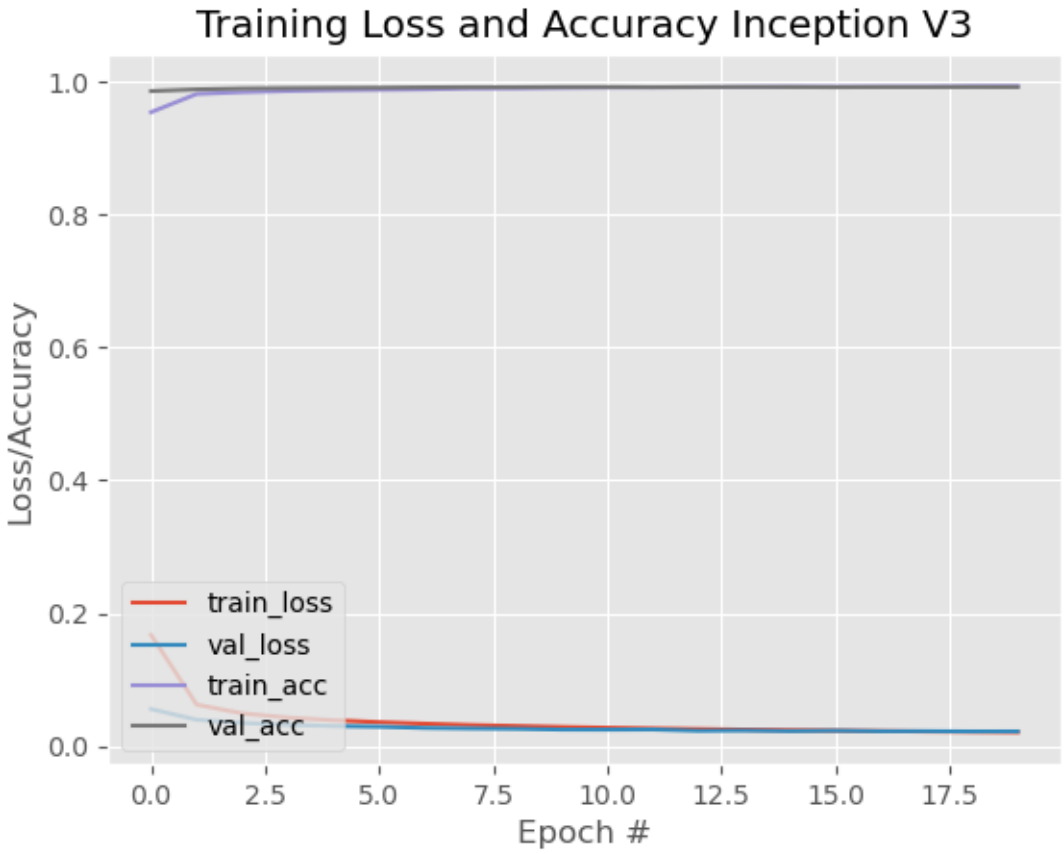


Fig.24 Inception V3 Training Graph.

5.3 Genes associated with Colorectal Cancer and Gene Expression

5.3.1 Introduction to Genetic Mutations in Colorectal Cancer

Genes control cell proliferation, apoptosis, differentiation, migration, and invasion, and imbalances in gene expression occur in a variety of human cancers. Recent studies have evaluated gene expression in colorectal cancer using microarray dataset or next-generation sequencing technology. The new technologies can help identify core genes related to the development of colorectal cancer, and these core genes are effective candidates for biomarkers in the development of colorectal cancer. Therefore, screening of core genes for colorectal cancer is very important. A list of core Up/Down regulated core genes is listed in **Table 7.**

Why should we test for Genetic Mutations?

Genetic tests can help show if members of certain families have inherited a high risk of colorectal cancer due to inherited cancer syndromes such as Lynch syndrome (also known as hereditary non-polyposis colorectal cancer, or HNPCC) or familial adenomatous polyposis (FAP).

In families known to have one of these inherited syndromes, family members who decide not to get tested are still usually advised to start routine screening for colorectal cancer at an early age, and to get screened more often. Family members who are tested and are found not to have the mutated gene may be able to be screened at the same age and frequency as people at average risk.

Testing for Lynch syndrome (hereditary non-polyposis colorectal cancer, or HNPCC)

Lynch syndrome: is the most common hereditary cancer syndrome. As many as 1 in 300 people carry a genetic mutation associated with Lynch syndrome and have a significantly increased risk of developing colon cancer. The lifetime risk of colorectal cancer in people with this condition can range from about 10% to 80%, depending on which gene mutation is causing the syndrome.

People with Lynch syndrome are also at increased risk for some other cancers, such as cancers of the uterus, ovaries, stomach, small bowel, pancreas, kidneys, brain, ureters (tubes that carry urine from the kidneys to the bladder), and bile duct.

Who should be tested for Lynch syndrome?

There are two sets of guidelines that doctors often use to determine who might be likely to benefit from genetic counseling or testing: the Amsterdam criteria (based on family history) and the revised Bethesda guidelines (for people diagnosed with colorectal cancer).

Amsterdam criteria

Doctors have found that many families with Lynch syndrome tend to have certain characteristics, which are known as the *Amsterdam criteria:*

At least 3 relatives have a cancer linked with Lynch syndrome and:

- One is a first-degree relative (parent, brother or sister, or child) of the other 2 relatives.

- At least 2 successive generations are affected.
- At least 1 relative had their cancer when they were younger than age 50.

If all of these apply to your family, then you might want to seek genetic counseling. But even if your family history satisfies the Amsterdam criteria, it doesn't always mean you have Lynch syndrome. And many families with Lynch syndrome do not meet the Amsterdam criteria.

Revised Bethesda guidelines

A second set of criteria, called the revised Bethesda guidelines, can be used to help decide whether a person with colorectal cancer should be tested for genetic changes that are seen with Lynch syndrome. These changes are called **microsatellite instability or MSI.** These criteria include at least one of the following:

- The person is younger than 50 years when diagnosed with colorectal cancer.
- The person has or had a second colorectal cancer, or another cancer (endometrial, stomach, pancreas, small intestine, ovary, kidney, brain, ureters, or bile duct) linked to Lynch syndrome.
- The person is younger than 60 years, and the cancer has certain characteristics seen with Lynch syndrome when it's viewed under a microscope.
- The person has a first-degree relative (parent, sibling, or child) younger than 50 who was diagnosed with colorectal cancer, or another cancer linked to Lynch syndrome.

- The person has 2 or more first or second-degree relatives (aunts, uncles, nieces, nephews, or grandparents) who had colorectal cancer or another Lynch syndrome-related cancer at any age.

Even if you don't have cancer, your doctor may suspect that Lynch syndrome runs in your family based on cases of colorectal cancer and other cancers associated with this syndrome in your relatives. In that case, your doctor might recommend genetic counseling to evaluate your risk.

The two major syndromes associated with a genetic tendency to colon cancer are: Lynch syndrome and Familial Adenomatous Polyposis syndrome (FAP)

- **Lynch syndrome,** which was discussed in the earlier section.
- **FAP Gene**: is characterized by more than 100 colorectal polyps in the colon. While it is less common – anywhere from 1 in 7,000 people to 1 in 22,000 have the condition, those with FAP have very high risks for develop colorectal cancer, at very young ages.
- **Other genes**: AXIN2, *BMPR1A*, *CDH1, CHEK2, GREM1, MUTYH, NTHL1, POLE, POLD1, PTEN, SMAD4, STK11, TP53, and others.*

If your family carries Lynch syndrome

In families known to carry a Lynch syndrome gene mutation, doctors recommend that family members who have tested positive for the mutation and those who have not been tested should start colonoscopy screening during their early 20s, or 2 to 5 years younger than the youngest person in the family with a diagnosis (whichever is earlier). Testing should be done every 1 or 2 years. This way polyps can be found and removed, and any cancers can be found at the earliest possible stage.

In summary, genetic testing is done with the support of colonoscopy polyps' detection and classification (which was discussed in earlier section) to ensure detecting the cancer in its earliest stage and take the suitable action on it to provide healthier population that have lower risk of getting colorectal cancer.

5.3.2 Dataset collection

We collected **66** colorectal cancer-associated genes from NCBI ***[33]***, and **97** (27, 70 up, and down regulated genes, respectively) core genes related to the development of colorectal cancer ***[34:36]***, which are listed in below.

Table 7. List of Genes associated with Colorectal Cancer.

Gene Types	Genes names
Genes associated with colon/colorectal cancer	AURKA, APC, AKT1, CCND1, BUB1B, AXIN2, DCC, BMPR1A, EPCAM, CTNNB1, EP300, BUB1, CHEK2, BRAF, BAX, IL1B, DLC1, KLF6, CASP10, FLCN, IRF1, MSH6, MCC, MLH3, FGFR3, EGFR, IL1RN, MLH1, KRAS, NRAS, PDGFRL, MSH2, PLA2G2A, PIK3CA, ERBB2, NTHL1, PTPRJ, RPS19, PTPN12, RABL3, SMAD4, RNF43, RAD54B, STK11, SRC, TLR2, TGFBR2, MUTYH, PMS1, ETV6, SMAD7, FGFR2, GALNT12, KLLN, UGT1A1, POLD1, POLE, SEC23B, POT1, MSH3, PMS2, TP53, SH3KBP1, PTEN, CDH1, GREM1
Up-Regulated Genes	CXCL1, CXCL3, CXCL8, DUSP14, SULF1, SOX9, TGFBI, CDC25B, FAP, COL4A1, MMP7, HILPDA, VSNL1, KRT23, MMP1, S100A2, ANXA3, UBE2C, DDIT4, AJUBA, EGFL6, COL10A1, COL11A1, MYC, TCF7, MET, LEF1
Down-Regulated Genes	SST, AQP8, CA7, SPIB, CA1, SLC4A4, GUCA2A, PYY, SCNN1B, GCG, HSD11B2, DHRS11, CA4, CLDN8, HSD17B2, ZG16, SLC26A3, ADAMDEC1, CA2, CHGA, LGALS2, LRRC19, TSPAN7, PTPRH, ABCG2, EDN3, CDHR5, HPGD, STMN2, CDHR2, MS4A12, CHP2, AKR1B10, ABCA8, TNFRSF17, ADH1B, SLC26A2, SLC30A10, MXI1, ENTPD5, CEACAM7, KRT20, CWH43, CLCA4, SELENBP1, SLC25A20, SULT1A2, ADH1C, NR3C2, DEFB1, AHCYL2, DHRS9, FABP4, CD177, CKB, CFD, ITM2A, TUBAL3, C7, LPAR1, DNASE1L3, GHR, RCAN2, ADTRP, SCN7A, BTNL8, PPP1R14D, FHL1, EPB41L3, EPHA7

5.3.3 Program Build

Firstly, we collected the genes normal protein sequences from NCBI ***[33]*** where every gene had one or more normal (Wild-Type) protein sequences, which resulted in **374** sequences.

Any change in the coding protein sequence of a gene will result in mutation in that gene which will lead to increased chance to cancer. With that knowledge, we decided to create a program that takes the protein sequence of a user and searches in our **374** normal protein sequences database if the entered sequence is normal (Wild-Type) or mutated.

5.3.3.1 How the Program Works

The user can input a protein coding sequence or a nucleotide sequence which can be translated in the program with the help of the amino acids wheel in **Fig.25**

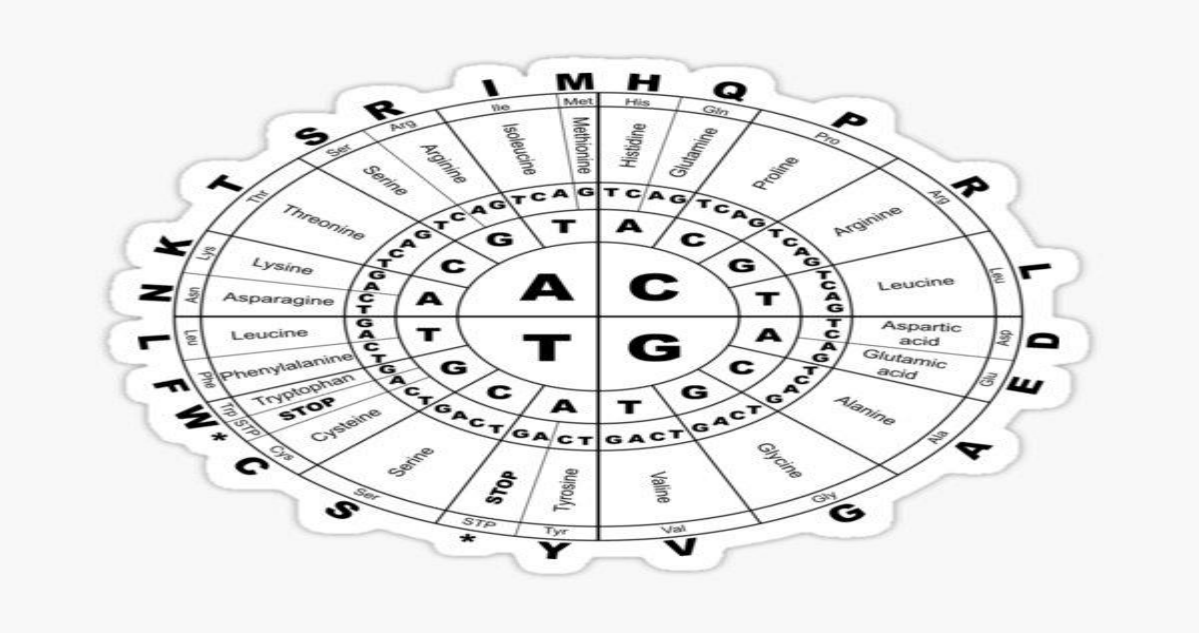


Fig.25 Amino Acid Wheel.

Recent studies show that in total, we all carry 100-200 new mutations in our DNA. This is equivalent to one mutation in each 15 to 30 million nucleotides. Fortunately, most of these are harmless and have no apparent effect on our health or appearance (Silent mutations). These silent mutations occur in the last nucleotide of a codon (three nucleotide sequences that codes for a certain protein), which almost all the time codes the same protein.

The previous amino acid wheel was programed to translate any nucleotide sequence to a protein sequence so that any nucleotide sequence entry can be translated correctly even if it had silent mutations in it, then we can accurately compare the entered protein sequence with our database to check for mutations.

This process is done by translation function which splits the nucleotide sequence into codons, then translates each codon by the programmed amino acid wheel database, and removing the last nucleotides if there are < 3. Finally, we have the translated protein sequence.

If the entered sequence was a protein sequence, the program skips the translation function and fasts forwards to the preprocessing and search function.

The preprocessing function ensures that the input sequence is coding sequence (exon) and is not non-coding sequence (intron) by removing any protein sequence before the start codon Methionine (M) and removing any protein sequence after any of the three stop codons (TAA, TAG, TGA) while removing any whitespaces in the entered sequence.

After translating the entered nucleotide sequence and preprocessing it, we search in our genes protein sequences database to find the most similar sequence for the entered one by computing the hamming distance normalized

similarity algorithm, with threshold of **30%** similarity/identity between the protein sequences. If the similarity is below **30%** according to research ***[37]***, then the program displays that the entered gene is not in our database.

While the similarity is above **30%**, the program returns the gene name, sequence, consensus coding sequence ID, URL, and gene type (Wild-Type or Mutated) of the gene that have the maximum similarity among all the genes and proceeds to the alignment function which aligns the entered gene sequence with the sequence of the maximum similarity gene and computes their pairwise scores with the following criteria:

- Match Score: 2
- Mismatch Score: -1
- Gap Opening Score: -0.5
- Gap Extension Score: -0.1

The alignment function also returns a graph of the pairwise sequence alignment to show where the mutation happened, if there are any. Also, the maximum score of which the sequences would have achieved if the entered gene sequence was normal (Wild-Type). Below is the flowchart of the program sequence in **Fig.26.**

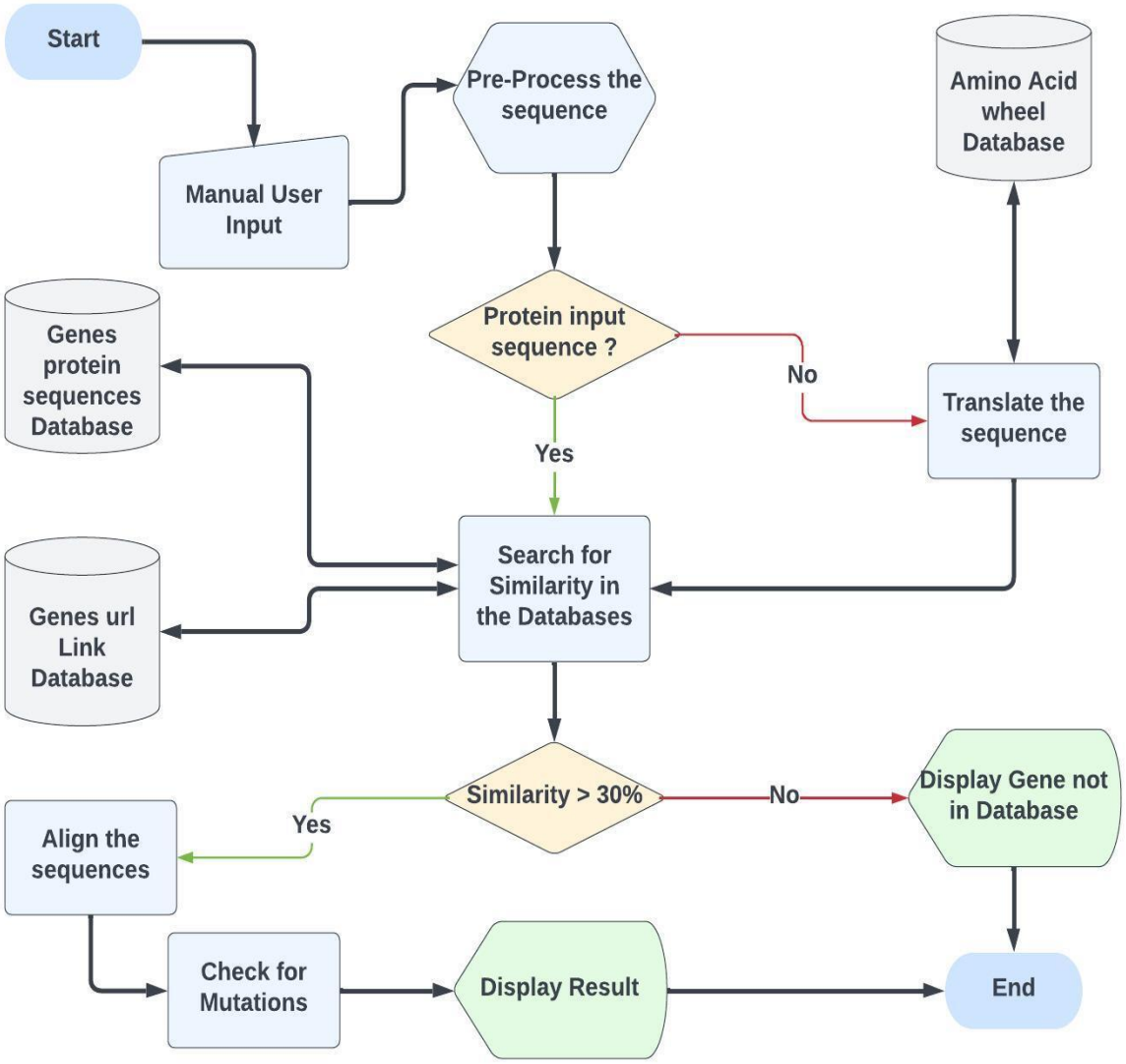


Fig.26 The Flowchart of the Program Sequence.

5.4 Colorectal Cancer Tumor Markers

Colorectal cancer (CRC) is a common type of cancer that affects the colon or rectum. Tumor markers are substances that are produced by cancer cells and can be detected in the blood or other body fluids. Tumor markers are often used in the diagnosis and management of CRC, as they can provide valuable information about the presence, progression, and treatment response of cancer.

The importance of tumor markers in CRC lies in their ability to provide additional diagnostic and prognostic information beyond what is obtained through imaging tests and biopsies. Tumor markers can help detect the presence of CRC at an earlier stage, monitor the progression of the cancer, and evaluate the response to treatment. Tumor markers can also be used to detect the recurrence of the cancer after treatment, which can help guide further treatment decisions.

One of the most used tumor markers in CRC is carcinoembryonic antigen (CEA). Elevated levels of CEA in the blood can indicate the presence of CRC or the recurrence of the cancer after treatment. CEA is also used to monitor the response to treatment and to detect any recurrence of the cancer. Other tumor markers that are used in the diagnosis and management of CRC include CA 19-9, CA 24-2, alpha-fetoprotein (AFP), and CA 50.

Here is some information about each of these tumor markers, including how to get tested for them, their importance, and their cut-off values:

Carcinoembryonic Antigen (CEA): CEA is a protein that is produced by normal fetal cells and is used as a tumor marker for CRC. Elevated levels of CEA in the blood can indicate the presence of CRC or the recurrence of the cancer after treatment. CEA is also used to monitor the response to treatment and to detect any recurrence of the cancer. The normal range for CEA is

generally considered to be less than **5 IU/mL**, although the cut-off value may vary depending on the laboratory.

CA 19-9: CA 19-9 is a carbohydrate antigen that is produced by some cancer cells, including those in the colon and rectum. Elevated levels of CA 19-9 in the blood can indicate the presence of CRC or the recurrence of the cancer after treatment. CA 19-9 is often used in combination with CEA to monitor the response to treatment and to detect any recurrence of the cancer. The normal range for CA 19-9 is generally considered to be less than **37 IU/mL**, although the cut-off value may vary depending on the laboratory and patient's age and gender.

CA 24-2: CA 24-2 is a carbohydrate antigen that is produced by some cancer cells, including those in the colon and rectum. Elevated levels of CA 24-2 in the blood can indicate the presence of CRC or the recurrence of the cancer after treatment. CA 24-2 is often used in combination with other tumor markers and diagnostic tests to evaluate CRC. The normal range for CA 24-2 is generally considered to be less than **20 IU/mL**, although the cut-off value may vary depending on the laboratory.

Alpha-Fetoprotein (AFP): AFP is a protein that is produced by the liver and is used as a tumor marker for liver cancer. Elevated levels of AFP in the blood can indicate the presence of liver cancer, although AFP may also be elevated in some cases of CRC. AFP is often used in combination with other diagnostic tests to evaluate CRC. The normal range for AFP is generally considered to be less than **15 IU/mL**, although the cut-off value may vary depending on the laboratory.

CA 50: CA 50 is a carbohydrate antigen that is produced by some cancer cells, including those in the colon and rectum. Elevated levels of CA 50 in the blood can indicate the presence of CRC or the recurrence of the cancer after treatment.

CA 50 is often used in combination with other tumor markers and diagnostic tests to evaluate CRC. The normal range for CA 50 is generally considered to be less than **25 IU/mL**, although the cut-off value may vary depending on the laboratory.

These cut off values are not fixed, they are dependent on the patient's age and gender, and we used the following research papers to get the most important Tumor Markers and their relative cut off values ***[38-39-40-41]***

To get tested for these tumor markers, a blood sample is typically taken and analyzed in a laboratory. The results of the test are then interpreted by a Healthcare provider, who may use the results to help diagnose and manage CRC.

The results of tumor marker tests can have a significant impact on the treatment of patients with CRC. For example, if a patient has elevated levels of CEA, their Healthcare provider may recommend additional diagnostic tests to evaluate the extent of the cancer and determine the best treatment options. If a patient has a recurrence of CRC after treatment, their Healthcare provider may use tumor markers to monitor the progression of the cancer and adjust their treatment plan accordingly.

In summary, tumor markers are an important tool in the diagnosis and management of CRC. While they should not be used as the sole diagnostic tool, they can provide valuable information that can guide treatment decisions and improve outcomes for patients with CRC.

5.5 Patients Tracking and Staging

After someone is diagnosed with colorectal cancer, doctors will try to figure out if it has spread, and if so, how far. This process is called *staging*. The stage of a cancer describes how much cancer is in the body. It helps determine how serious the cancer is and how best to treat it. Doctors also use a cancer's stage when talking about survival statistics.

The earliest stage colorectal cancers are called stage 0 (a very early cancer), and then range from stages I (1) through IV (4). As a rule, the lower the number, the less the cancer has spread. A higher number, such as stage IV, means cancer has spread more. And within a stage, an earlier letter means a lower stage. Although each person’s cancer experience is unique, cancers with similar stages tend to have a similar outlook and are often treated in much the same way.

5.5.1 Determination of the Stage *[42]*:

5.5.1.1 TNM Staging

The staging system most often used for colorectal cancer is the American Joint Committee on Cancer (**AJCC**) **TNM** system, which is based on 3 key pieces of information:

The extent (size) of the **t**umor **(T):** How far has the cancer grown into the wall of the colon or rectum.

The spread to nearby lymph **n**odes **(N):** Has the cancer spread to nearby lymph nodes.

The spread (**m**etastasis) to distant sites **(M):** Has the cancer spread to distant lymph nodes or distant organs such as the liver or lungs.

Numbers or letters after T, N, and M provide more details about each of these factors. Higher numbers mean the cancer is more advanced. Once a person's T, N, and M categories have been determined, this information is combined in a process called *stage grouping* to assign an overall stage.

5.5.1.1.1 Stages Grouping and Description

Table 8. A Summary of each AJCC Stage and Stages Grouping:

AJCC Stage	Tumor (T)	Node (N)	Metastasis (M)
0	Tis	N0	M0
	T0	N0	M0
I	T1 or T2	N0	M0
	TX	NX	M0
IIA	T3	N0	M0
	TX	NX	M0
IIB	T4a	N0	M0
	TX	NX	M0
IIC	T4b	N0	M0
	TX	NX	M0
IIIA	T1 or T2	N1 or N1c	M0
	T1	N2a	M0
	TX	NX	M0
IIIB	T3 or T4a	N1 or N1c	M0
	T2 or T3	N2a	M0
	T1 or T2	N2b	M0
	TX	NX	M0
IIIC	T4a	N2a	M0
	T3 or T4a	N2b	M0
	T4b	N1 or N2	M0
	TX	NX	M0
IVA	Any T	Any N	M1a
	TX	NX	
IVB	Any T	Any N	M1b
	TX	NX	
IVC	Any T	Any N	M1c
	TX	NX	

5.5.1.1.2 Description:

0: The cancer is in its earliest stage. This stage is also known as carcinoma in situ or intramucosal carcinoma (Tis). It has not grown beyond the inner layer (mucosa) of the colon or rectum.

Or there is no evidence of a primary tumor.

I: The cancer has grown through the muscularis mucosa into the submucosa (T1), and it may also have grown into the muscularis propria (T2). It has not spread to nearby lymph nodes (N0) or to distant sites (M0).

IIA: The cancer has grown into the outermost layers of the colon or rectum but has not gone through them (T3). It has not reached nearby organs. It has not spread to nearby lymph nodes (N0) or to distant sites (M0).

IIB: The cancer has grown through the wall of the colon or rectum but has not grown into other nearby tissues or organs (T4a). It has not yet spread to nearby lymph nodes (N0) or to distant sites (M0).

IIC: The cancer has grown through the wall of the colon or rectum and is attached to or has grown into other nearby tissues or organs (T4b). It has not yet spread to nearby lymph nodes (N0) or to distant sites (M0).

IIIA: The cancer has grown through the mucosa into the submucosa (T1), and it may also have grown into the muscularis propria (T2). It has spread to 1 to 3 nearby lymph nodes (N1) or into areas of fat near the lymph nodes but not the nodes themselves (N1c). It has not spread to distant sites (M0).

Or The cancer has grown through the mucosa into the submucosa (T1). It has spread to 4 to 6 nearby lymph nodes (N2a). It has not spread to distant sites (M0).

IIIB: The cancer has grown into the outermost layers of the colon or rectum (T3) or through the visceral peritoneum (T4a) but has not reached nearby organs. It has spread to 1 to 3 nearby lymph nodes (N1a or N1b) or into areas of fat near the lymph nodes but not the nodes themselves (N1c). It has not spread to distant sites (M0).

Or The cancer has grown into the muscularis propria (T2) or into the outermost layers of the colon or rectum (T3). It has spread to 4 to 6 nearby lymph nodes (N2a). It has not spread to distant sites (M0).

Or The cancer has grown through the mucosa into the submucosa (T1), and it might also have grown into the muscularis propria (T2). It has spread to 7 or more nearby lymph nodes (N2b). It has not spread to distant sites (M0).

IIIC: The cancer has grown through the wall of the colon or rectum (including the visceral peritoneum) but has not reached nearby organs (T4a). It has spread to 4 to 6 nearby lymph nodes (N2a). It has not spread to distant sites (M0).

Or The cancer has grown into the outermost layers of the colon or rectum (T3) or through the visceral peritoneum (T4a) but has not reached nearby organs. It has spread to 7 or more nearby lymph nodes (N2b). It has not spread to distant sites (M0).

Or The cancer has grown through the wall of the colon or rectum and is attached to or has grown into other nearby tissues or organs (T4b). It has spread to at least one nearby lymph node or into areas of fat near the lymph nodes (N1 or N2). It has not spread to distant sites (M0).

IVA: The cancer may or may not have grown through the wall of the colon or rectum (Any T). It might or might not have spread to nearby lymph nodes. (Any N). It has spread to 1 distant organ (such as the liver or lung) or distant set of

lymph nodes, but not to distant parts of the peritoneum (the lining of the abdominal cavity) (M1a).

IVB: The cancer might or might not have grown through the wall of the colon or rectum (Any T). It might or might not have spread to nearby lymph nodes (Any N). It has spread to more than 1 distant organ (such as the liver or lung) or distant set of lymph nodes, but not to distant parts of the peritoneum (the lining of the abdominal cavity) (M1b).

IVC: The cancer might or might not have grown through the wall of the colon or rectum (Any T). It might or might not have spread to nearby lymph nodes (Any N). It has spread to distant parts of the peritoneum (the lining of the abdominal cavity) and may or may not have spread to distant organs or lymph nodes (M1c).

If the main tumor cannot be assessed due to lack of information, then (TX), and if regional lymph nodes cannot be assessed due to lack of information, then (NX).

5.5.1.2 Grade (G) Staging

Doctors also describe colorectal cancer by its grade (G). The grade describes how much cancer cells look like healthy cells when viewed under a microscope.

The doctor compares the cancerous tissue with healthy tissue. Healthy tissue usually contains many different types of cells grouped together. If the cancer looks like healthy tissue and has different cell groupings, it is called "differentiated" or a "low-grade tumor." If the cancerous tissue looks very different from healthy tissue, it is called "poorly differentiated" or a "high-grade tumor." The cancer’s grade may help the doctor predict how quickly the cancer will spread. In general, the lower the tumor’s grade, the better the prognosis.

5.5.1.2.1 Stages Description

- **GX:** The tumor grade cannot be identified.
- **G1:** The cells are more like healthy cells, called well differentiated.
- **G2:** The cells are somewhat like healthy cells, called moderately differentiated.
- **G3:** The cells look less like healthy cells, called poorly differentiated.
- **G4:** The cells barely look like healthy cells, called undifferentiated.

In summary, we used the above staging methods alongside tumor markers to track the patient treatment course schedule and to analyze how the treatment affect the patient for faster recovery if the cancer is detected throughout our project methods.

5.6 Patients Statistics

5.6.1 Dataset Collection

We collected some colon cancer patient information records from Kafr El-Sheikh Oncology Institute under the supervision of **Prof. Emad Sadaka** and some other not-colon-cancer-related patient information records for analysis and visualization.

The dataset consists of **138** colon cancer patient records and **90** other patient records.

5.6.2 Methods Used

We used Python for analysis and Power BI tool for visualization purposes.

5.6.3 Conclusion

The analysis and visualizations draw our attention to these points:

- There’s a strong connection between overweight patients and the chance of having colorectal cancer.
- The chance of women who are smokers to get colorectal cancer is higher than in men.
- The chance of overweight men to get the cancer is higher than women.
- The average age of colorectal cancer patients is significantly lower than other patients.