[[1]](#footnote-1)

Convolutional Neural Network for

Polyps classification.(April 2022)

3724159 IEEE.

*Abstract*—The primary cause for the deaths of variety of diseases is usually a delayed diagnosis. There are numerous variety of diseases that cause fatalities. Polyps is one of those diseases that are usually diagnosed in a late stage, and is the reason for colorectal cancer. This cancer type is the third most common in USA, and it has caused 52,980 death in 2021. Some symptoms such as synopsis which make a person feel pain in stomach are not taken seriously and people try to avoid doctor consult for them. Patients are sometimes scared of their symptoms, and for this reason it is needed a solution that can help people identify the symptoms of a disease. Till now, gastroenterologist have had the major role for their classification, based on their appearance, size, form, and etc. However, the human being can sometimes do mistakes. This is the reason why in this article, there is proposed the Convolutional Neural Network classification for polyp types.

*Index Terms*— Benign, Cancer prediction, Machine Learning, Convolutional Neural Network, Polyp classification.

**Abbreviations and Acronymous:**

**ML- Machine Learning**

**USA- United States of America**

**CNN- Convolutional Neural Networking**

# INTRODUCTION

P

olyp is a small clump of cells which develop in the inner linings of the colon. Usually polyps are harmless unless it is not treated for a longer time period. After that, it might change into a fatal disease that is colon cancer. There are many precautions and prevention that need to be taken in the earlier stages for polyps not to increase, such as maintaining healthy diet, not eating fatty food, red meat, maintaining healthy life style.

Polyps Detection is a method that checks if a polyps at colon is Malignant or not. After giving its prediction this method advice people to ask for medical help. This method can be used in hospital for people with colon cancer about giving them a prediction about their future treatment, and doctor expectations.

The aim of this project is to develop a classification system for polyps that provides a solution to the doctors for polyps classification as malignant or benign. A system can analyses the colon image, suggest a classification, and next stages of treatment. Our proposed system tries to find the early symptoms of polyps in cell imaging when it is medical. The proposed solution aims to provide a pre-evaluation for this issue of polyps. This article is divided into 6 main sections.

First section talks about polyps description and shows the construction of colon layers which are mucosa submucosa, muscularis propria, and serosa. It also states the colon polyps which are growth in the colon that are mostly common at people over 40 years. Polyps can be seen with microscope, and they are classified into 4 different types which are Adenomas, Hyperplastic, Hamartomas, and Inflammatory. Each type has its own shape, size, and different risk rate for causing cancer. Polyps are mostly generated from parent to child, and some of its symptoms are rectal bleeding, loss of weight, change of bowel function, and etc. Furthermore, in this section is talked about colonoscopy that is the most popular method to inspect a column, and find polyps inside of it. After a colonoscopy, the chances for a polyp to repeat are between 20%-30%.

Second section talks about some facts of colon cancer in USA. In 2021 there were 149,500 new cases diagnosed. For people over 50 there is a small increase by 2% of polyps. Death rates has dropped by 55% from 1980 to 2018 because of better health system, and polyps detection. In this section it is stated that polyps detection on early stage can increase lifesaving rates by 91%.

Third section talks about the MI\_dataset provided by LSBU to work on polyps classification. This dataset is made of 207 records of polyps images. There are 4 types of polyps: Adenomas, Serrated Adenomas, Hyperplastic, and Adenocarcinoma. These polyps are classified in 2 classes malignant(45 instances), and benign(162 instances). There were no outlier detected in the data preprocessing of this dataset.

The fourth section talks about some previous work done for polyps classification. These works have used videos of endoscopy. These videos have been divided into many sequential images. They have used CNN method to predict the classification of polyps. This method has given some accurate results for classification between 74% and 78% accuracy. However, because of the small number of instances in dataset the accuracy has not been in its higher performance. Our model will try to bring a higher accuracy for polyps classification.

The fifth section talks about Naïve Bayes classification model used for the dataset analyze, and polyps classification. This model is very useful in small datasets and offer a good accuracy. The dataset is divided into 20% used for testing, and 80% used for training. The output of this algorithm was 78% accuracy which is satisfactory result.

Fifth section discussed about the methodology used to train the data of the polyps image dataset, and use of Transfer Learning, and CNN technique to train the dataset and give a certain prediction about polyps classification. The pictures of the dataset were transformed in order to increase the capacity of the dataset, and give a higher accurate result.

The last section discuss the accuracy of the model which was around 75% and express the future vision further researches that can be done to improve the result and to bring a higher accuracy. Some techniques are deep learning, time series, and AI which will give a higher accuracy.

# Colon polyps

Polyps of colons are growth in inner part of it. There are more commonly in people over 40. Colon is an important part fo digestive system that works with other vital organs. Colon is separated into 4 layers which can be affected by cances that starts at mucosa, and continues in other inner layers. there are 4 layers of the colon which have specific function [1]. The 4 layers of colon are:

1. Mucosa where many diseases happen and play the role of lubricant.
2. Submucosa which is made of vessels/nerves and is connected to muscles.
3. Muscularis propria which is constructed by fibres of muscles.
4. Serosa outer layer of the colon.

## Symptoms of colon polyps

Polyps are often a condition transmitted from parent to children. Sometimes they are often caused by unhealthy lifestyle such as smoking or alcohol. People at age over 50 are more in risk than other for this disease.

Polyps do not usually show any symptoms, and people can have them for many years without knowing. However Rectal bleeding, or abdominal pain can sometimes be symptoms of colon polyps. Some people may also lose weight in a short time [2].

## Colonoscopy

Colonoscopy is the most common used technique to spot a polyps in a colon. This method checks the colon from inside, and takes a part of polyps to be analyzed, They can be in different size, and different parts of colon. Figure 1 shows some polyps types that are in different sizes. These polyps are in different stages, and if they are not treated at an early stage they can become cancer. The time needed for a polyp to become cancerous are many year but it can be very quick a life risk if it touches other vital organs.

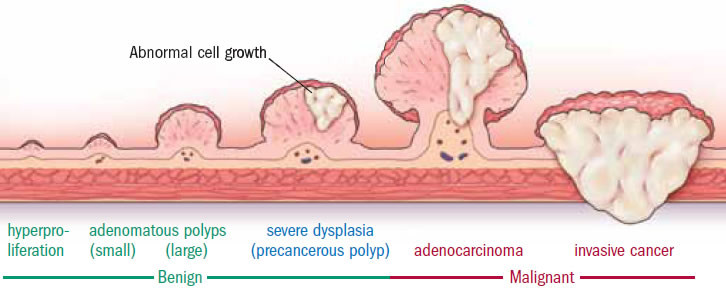


Figure 1- Colon polyp progress. Harvard Health publishing, “They found colon polyps: Now what?” 10 September, 2019. <https://libraryguides.vu.edu.au/ieeereferencing/figurestablesequations>

20% to 30% are the chances for a polyps to be repeated if its removed [3]. If a Colonoscopy finds small polyps with 5mm diameter there is a low risk for cancer. The risk for a cancer polyps increases if it has a diameter bigger than 10 mm, and does not have a normal look.

# Statistic of Colon Cancer

As far as we know, the Colon cancer, and colorectal cancer are one of the main diseases that are caused by the column polyps that might cause deaths and very common in cancer diagnosis. According Cancer.net, colorectal cancer takes the third place as the most common cancer type in the United States. Around 150 000 new cases were diagnosed with this diseases in 2021 where around 104 000 were new cases which are reported as colon cancer. The cases of colorectal cancer has dropped by 1% each year from 2013 to 2017 [4]. However there is a slightly increase of the cases by 2% of this cancer for adults whose age is under 50, and 1% slight increase for adults between 50 to 64

The deaths caused by colorectal cancer, and colon cancer is around 53 thousand which is the second cause of cancer death in the United States.

The Colorectal, and colon cancer should be found early in order to be cured. Death rate has dropped exponentially by 55% compared between 2018, and 1970 In US. The reasons are the improvements in health treatment, and the increase of screening the cancer cells at an early stage. The death rates has dropped slightly , around 2% from 2014 to 2018. However, there are around 1.5 million people diagnosed with colorectal cancer.

Surgical removal of tumors can improve 5-years survival rates of the patients, and can eliminate the cancer. The survival rate after the surgical removal has increased by 63% for people of 5-year survival. If Colon cancer is diagnosed and localized early, it can save the life up to 91%. If colon cancer has affected other parts of body, there are around 14% chances of surviving in the next 5 years.

As it can be seen, the diagnosis of cancer at early stage can increase the chances of patients to survive. This is the main reason why in these study it is being worked on the colon polyps classification so people with high risk of cancer can be advised to take a surgery, and medical help. According to National Polyp Study shows that a colonoscopy at initial stage can reduce the risk of cancer by 90%, and it can save lives from dying up to 92% [5].

## Polyps types

The polyps identified in “MI\_polyps” dataset are divided into 4 types. These types are adenomas, Serrated Adenomas, Hyperplastic, and Adenocarcinoma.

Adenomas polyps are usually harmless, and can be precancerous. If they are not removed, they can grow and develop cancer. They can be grown in mucous membrane [6]. Tubular, and villous are two common types of adenomas. Tubular adenomas are usually in small size with tubular shape, and are usually less than half inch long. Villous adenomas is finger-leaf coming out adenoma.

Serrated Adenomas have “sawtooth” pattern on polyps surface [7]. This polyp type is usually harmless, and can be cured if it is removed at an early stage. There are 2 types of serrated adenomas that look different where seen in microscope. They are named Sessile serrated adenomas, and Traditional serrated adenomas.

Hyperplastic polyp is a growth in areas of the colon. They are usually benign if they are not in big sizes [8]. However there are some cases where hyperplastic polyps are malignant. According to the shape there are 3 types of hyperplastic polyps. First type is pedunculated that are long, and narrow. Second types is sessile that are short, and look as squat. The third type is serrated that is flat, wide at the bottom, and short.

Adenocarcinoma polyps is composed of well-formed glands with open lumina [9]. If this polyp is not diagnosed at an early stage there is a higher risk of being malignant, and have higher chances of causing cancer. Adenocarcinoma can start in cells , and make mucus lubricate inside the colon. It is divided into 4 stages that can cause big damages on the human body.

# Related works for polyp classifications

In other Studies, different approaches are used previously for Polyp classification and in all these studies that are made are used data from colonoscopy. During recent studies, polyps classification studies did not give a real high accuracy classification and one reason is that the dataset is small, which leads to insufficiency for training a model thus giving an error rate close to 0%. Some previous studies which were published on Pone journal used a dataset that contains 157 polyps videos and they were split into many images 27,048. This sequence of images was split into 2 groups, hyperplastic and adenomatous. With the help of endoscopists from medical school Jilin University on what polyps classification they had to use to get high overall accuracy, they achieved an overall accuracy of 75.71% and 79.78% for Set 1 and 2 respectively [11]. The overall accuracy was not bad however their aim was to get a higher rate so they tried using ACU ( Area under the curve) and RCO (Receiver Operation Characteristics) to receive a higher overall performance than the first tries. They used about 6 deep learning classification models for result comparison like DenseNet, MasNet, ResNet, SENet but did not perform very well on test. However, they gave few ideas about what classification should be used in the future researches like an algorithm that might work better on small datasets since there are not many records on Polyps Dataset. In 2018 there was another study which was posted in Gastroenterology that shows use of CNN (Deep convolutional neural networks to increase detection rate of Adenoma (ADR). On this study, 8641 images of colonoscopies were analyzed for 4088 unique polyps and tested five hour videos of 20 different colonoscopy. Expert could identify around 9 videos of polyps. They used CNN for testing these videos and this algorithm gave an accuracy of 96.4% by finding polyps under receiver operating system curve [12]. 28 Polyps are removed and identified by experts and detected by Neural Networks algorithm, with 7% as false-positive ratio. On this study, they provided polyps detection on real time and on a regular desktop machine. However, Polyps type was not classified on this study and a large multicenter trials validation is required.

# Methodology

"Transfer Learning" is one way for analyzing large datasets. This strategy essentially involves reusing a previously learned model. Because deep neural networks can be trained on little datasets, this approach has become increasingly popular in deep learning. We'll look at how transfer learning works, what it is, why it's important, and when it's appropriate to apply it. The early and intermediate layers are utilized during Transfer Learning, but the latter layers are really retrained. It aids in the utilization of the task's labelled data when it is first learned. Medical imaging might benefit from this technology in the future. ConvNets allow machines to see the environment in the same way that people do, allowing them to do tasks such as Image Video Recognition, Media Recreation, Image Analysis Classification, and Recommendation Systems, among others. The connection network of Nerve cells in the Human Nervous System is comparable to and related to the design of a ConvNet, which is impacted by Visual Cortex. Changes in tiny regions of receptive acuity, known as the Receptive Field, are all that neural circuits can react to. To fill up the whole peripheral vision, similar fields might be piled on top of each other.

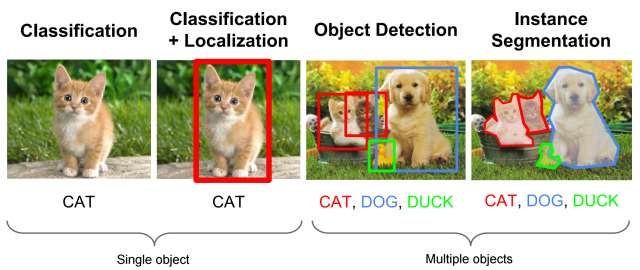


Figure 3. An example of Convolutional Neural Network

## Data Information

Professor Enrico Grisan of London South Bank University in London, United Kingdom, has given the polyps dataset. This collection contains 207 photos of polyps from two distinct polyp kinds. Malignant class "1" polyps are classed as malignant, whereas benign class "0" polyps are classified as benign. The dataset was a significant challenge because it was impossible to locate one that was similar, despite the fact that the internet has a wealth of datasets for various machine learning applications. Some files were left unprocessed. Data pictures in.mat formats were used for this study. CSV files have also included metadata.This collection includes images of benign and malignant polyps. The model calculates the likelihood as a number between 0 and 1. The class names themselves have been used to signify benign in the train datasets. The training data for the model is contained in the csv file. The data in the "fp" column includes the unique filename of each image. The term "diagnostic" refers to the sort of diagnosis used in imaging. This column, "Type," displays the indication of imaging malignancy.

Graphical user interface, text

Description automatically generated

Figure 4- Sample Dataset B. Data Retrieval

Because the photos were in.mat format, they required to be transformed to a CNN-readable format, such as png or jpeg, before being input into the classification models. We used the scipy.io loadmat method to load the mat file into an array, then converted it to an image using PIL's image from array function. We discovered that there are numerous unclean photos that are either colorless or have more than three dimensions or fewer than two dimensions during this procedure. As a result, after we discovered the issue, we filtered out the dataset including those photographs. The information was gathered and organized into railway folders. Then, for Data Retrieval, it utilized the Keras image generator from the Keras package to partition the training set into train and validation sets, with the validation set having a 0.2. In addition, we utilized the rescaling factor 1./255 for each of the photos received to ensure that no information from the photographs was lost. We kept the picture size that would be retrieved at (224, 224, 3) and utilized the batch size of 64, which we discovered to be the most efficient. However, we discovered that the collection contains extremely few photographs.. As a result, in order to avoid the problem of overfitting, we opted to enrich the dataset using Keras Data Augmenter, which we did using the following techniques: • Make a 90-degree turn. • Make a 270-degree turn. • Make a left-to-right flip. • Make a top-to-bottom flip. • A random crop with a 0.5 percent area. • Image resizing C. On this project, we primarily employed three models for data modelling, with InceptionV3 and Resnet50 serving as two of the models. Because it is built for this, GoogleNet's inception architecture operates admirably even when memory and computational resources are limited. Changes to the network are more difficult due to the network's intricacy. The conception style building blocks have a flexible structure that allows them to automatically absorb restrictions. This is made possible by the inception modules' extensive use of dimensional reduction and parallel structures, which reduce the impact of structural alterations on surrounding components. ResNet was named the winner of the ImageNet Challenge in 2015, and it was the first model to incorporate the notion of skip connection. Inception, also known as GoogleNet, has 152 levels and was used to train incredibly deep networks with over 150 layers. Aside from these pre-trained networks, we created a bespoke model utilising keras' Sequential features. We used numerous blocks, each with two convolutional layers and a single MaxPool layer with Relu as the activation function and valid padding, as well as a stride of one size and 32 filters. Five of these layers were utilized, followed by a batch normalization layer. The findings are then flattened, with numerous thick layers and a single dropout layer following. The custom model can be seen in Figure 3. D. Presentation and Results Following that, we trained our dataset on each of the aforementioned models independently. The findings of the inception V3 model may be shown below in Conclusion.

The data obtained from the inception V3 model clearly demonstrate that train accuracy improves with time, but validation remains constant, if not slightly decreasing, and the worst thing to note is that validation loss rises over time, indicating that this model is not suitable for usage. Our unique model, while similar to the Inception V3 model, has a falling validation accuracy and a peaked validation loss, making it unsuitable for application. Finally, Resnet is the final model.. We can clearly see that the validation accuracy either decreases or increases, which is better than the above two. We can also see that the validation and training loss diminishes with time, leading us to believe that this model is neither overfit nor underfit. As a result, with a 75 percent accuracy, is a solid model to utilize.

# Conclusion

In deep learning classification problems there are a lot of metrics that we have to take care of in order to get good results.

Graphical user interface

Description automatically generated

Fig. 3. Custom Model Details

Amongst which controlling the under-fitting and over fitting is a major one, for which we used dropout layers and increasing the dataset size by using augmentation techniques to tackle these problems. We also trained our models on pre-trained models using transfer learning and also used our custom model to use for our purpose. After observing all the results obtained and graphs made from the models, we can clearly state that RESNET50 performed overall better that all the models that we have used by providing us with a sound accuracy of 75%.

# FUTURE WORK

This model can be further studied using Time Series model that can use data available in hospitals to predict the cases of polyps cancer, and their possible cause. In conclusion, usage of AI-augmented digital system could help to increase the accuracy score compared to microscopic examinations has an error rate around 10% higher than AI-augmented digital system [10]. In our classification system we achieved a satisfactory score around 80% which seems to be higher than most microscopic examination’s score 74%. However there are some limitations on using digital system like the location of biopsy, which seems to be vital attribute for classification so the best practice is to combine both the Clinical Metadata and the Polyp images so we can have a higher success rate. The usage of Artificial Intelligence or also called Deep Machine Learning in clinical practice might be associated with limited overuse or underuse of subsequent surveillance colonoscopy, reduced stress, financial burden and most importantly, prevention of catching the diagnosis of cancer too late which may be fatal to a human.

Appendix and the use of Supplemental files

# References

|  |  |
| --- | --- |
| [1] | C. Vincent, "The Function of the Colon," 28 October 2021. [Online]. Available: https://www.verywellhealth.com/what-is-the-colon-796819. [Accessed 10 January 2022]. |
| [2] | M. clinic, "Colon polyps," 20 July 2021. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/colon-polyps/symptoms-causes/syc-20352875. [Accessed 02 01 2022]. |
| [3] | H. H. Publishing, "They found colon polyps: Now what?," 10 September 2019. [Online]. Available: https://www.health.harvard.edu/diseases-and-conditions/they-found-colon-polyps-now-what. [Accessed 29 12 2021]. |
| [4] | Cancer.net, "Colorectal Cancer: Statistics," 01 2021. [Online]. Available: https://www.cancer.net/cancer-types/colorectal-cancer/statistics. [Accessed 13 01 2022]. |
| [5] | L. Chang, "Removing Colon Polyps Cuts Death Risk," 15 October 2007. [Online]. Available: https://www.webmd.com/colorectal-cancer/news/20071015/removing-colon-polyps-cuts-death-risk. [Accessed 29 11 2021]. |
| [6] | A. Rowden, "What to know about adenomas," 29 March 2021. [Online]. Available: https://www.medicalnewstoday.com/articles/adenomatous-polyps#definition. [Accessed 14 12 2021]. |
| [7] | J.-Q. J. B.-S. D. G.-G. J. Monreal-Robles R, "Serrated polyps of the colon and rectum: a concise review.," 9 June 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34116964/. [Accessed 15 11 2021]. |
| [8] | T. Jewell, "Everything You Need to Know About Hyperplastic Polyps," 18 September 2018. [Online]. Available: https://www.healthline.com/health/hyperplastic-polyp#colon. [Accessed 03 12 2021]. |
| [9] | A. C. I. G. a. J. I. A.-M. Luis Bujanda, "Malignant colorectal polyps," 07 July 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896747/. [Accessed 17 12 2021]. |
| [10] | K. L. T. W. B. R. W. Krushi Patel, "A comparative study on polyp classification using convolutional neural networks," *Plos One,* vol. I, no. 30, p. 8, 2020. |
| [11] | P. T. T. A. M. M. F. J. W. K. P. B. Gregor Urban, "Deep Learning Localizes and Identifies Polyps in Real Time with 96% Accuracy in Screening Colonoscopy," *Gastroenterology,* vol. 4, no. 8, pp. 1069-1978, 2018. |
| [12] | C. Murakami, "Polyps in the Large Intestine (Colon): Types, Causes, Risk Factors, and Diagnosis," 11 October 2021. [Online]. Available: https://www.emedihealth.com/digestive-health/health-conditions/manage-colon-polyps#:~:text=Types%20of%20Colon%20Polyps%20There%20are%20four%20different,into%20cancer%20over%20several%20years%20if%20not%20removed.. [Accessed 12 1 2022]. |
| [13] | A. Gonfalonieri, "5 Ways to Deal with the Lack of Data in Machine Learning," 10 June 2019. [Online]. Available: https://www.kdnuggets.com/2019/06/5-ways-lack-data-machine-learning.html. [Accessed 11 01 2022]. |
| [14] | S. A. Mohammad Ali Kadampur, "Skin cancer detection: Applying a deep learning," *Informatics in Medicine,* vol. 18, no. 100282, pp. 2352-9148, 2020. |
| [15] | L. K. T. K. W. Q. B. A. R. Patel K, "A comparative study on polyp classification," *PLoS ONE,* p. 15, 2020. |
| [16] | U. A. W. G. H. M. Ribeiro E, "Exploring Deep Learning and Transfer Learning for," *Comput Math Methods Med,* 2016. |
| [17] | A. M. O. A. P. M. C. M. N. M. A. S. L. T. A. A. S. S. H. Bruno Korbar1, "Deep learning for classification of colorectal," 25 July 2017. [Online]. Available: https://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2017;volume=8;issue=1;spage=30;epage=30;aulast=Korbar. [Accessed 18 February 2022]. |

## 

1. [↑](#footnote-ref-1)