جـــامـعـــة الـــشـرق الأوسـط
MIDDLE EAST UNIVERSITY
Amman - Jordan

DeepCPC: Deep Learning Model for Colorectal Polyps Classification

DeepCPC: نموذج التعلم العميق لتصنيف سلائل أورام القولون والمستقيم

Prepared by:
Dima Hussein Taha

Supervisor:
Dr. Ahmad Gazi Al zu'bi

Submitted in Partial Fulfilment of the Requirements of the Master Degree in Computer Science

Computer Science Department
Faculty of Information Technology
Middle East University
May, 2020

Authorization

I, **Dima Hussein Fakhri Taha**, authorize Middle East University to provide Libraries, Organization, and Individuals with copies of my thesis on request.

Name: Dima Hussein Fakhri Taha.

Date: 09/06/2020.

Signatures:

Examination Committee Decision

This is to certify that the thesis entitled "DeepCPC: Deep Learning Model for Colorectal Polyps Classification" was successfully defended and approved on 28/5/2020.

Examination Committee Members	Signature
(Supervisor) **Dr. Ahmad Alzu'bi** Assistant Professor, *Department of Computer Science* *Middle East University*	
(Chairman of Examination Committee and Internal Committee Member) **Dr. Abdelrahman Abuarqoub** *Assistant Professor, Department of Computer Science* *Middle East University*	
(External Committee Member) **Dr. Adnan Hanif** *Associate Professor, Department of Computer Science* *Al-Zaytoonah University of Jordan*	

Dedication

*To My **Mother** Soul…..*

To My Sister

Esra'a Taha

For your endless love and support.

To My Husband

Said Al-Alami

Who inspired me all the time & always there for me.

To My Beloved Children

Maya & Adam

You deserve the best I have to give.

Acknowledgement

First and foremost, I would like to take this opportunity to express my profound gratitude and deep regard to my supervisor Dr. Ahmad Alzu'bi for his exemplary guidance, monitoring, invaluable constructive criticism, and friendly advice and constant encouragement throughout this thesis.

Furthermore, I would like to express a deep sense of gratitude to the dean Dr. Abdelrahman Abuarqoub, Dr. Hisham Abu-Saymeh, Prof. Hamza Al-Sewadi, and Dr. Mudhafar Al-Jarrah, for their cordial support, valuable information, and extending all the guidance during my master's studies period.

Lastly, I would like to thank all my friends, supporting me during my hard times.

Without your assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished.

DeepCPC: Deep Learning Model for Colorectal Polyps Classification

Prepared by

Dima Hussein Taha

Supervised By
Dr. Ahmad Gazi Al zu'bi

Abstract

Colorectal cancer is a silent disease that attacks without warning, but in many cases, treatment is possible when discovered early. Screening tests play an important role in identifying polyps before they become cancerous, where colonoscopy is more effective compared to other tests. Over the past few decades, the computer-aided colorectal polyp in colonoscopy has been the subject of research and achieved significant advances. However, the automatic polyp classification in real-time is still a challenging problem due to utilizing the hand-crafted methods that do not provide discriminating image features.

The advanced deep convolutional neural networks (DCNN) have shown a significant revolution that positively influenced many fields including image classification. In the domain of colonoscopy images, many limitations could affect the DCNN-based polyp's classification especially the lack of sufficient amount of patients' training samples, inadequate training time, and needed resources for neural networks.

The work in this thesis aims to develop a deep learning model for classifying colorectal polyps (referred to as DeepCPC), based on discriminative features extracted from deep conventional neural networks. Specifically, some CNN models pretrained has been used on general-purpose images to apply a transfer learning scheme in the polyp's classification system. This is achieved by concatenating a set of discriminating image features extracted from the activations of convolutional layers, then improved feature representations by fine-tuning a proposed CNN architecture on polyps images through a complete end-to-end training procedure. The proposed model consists of three main components: lower fine-

tuned layers, concatenated image vector, and fully-connected top layers. The CVC-ClinicDB dataset is used to evaluate the Deep CPC model, but further patch extraction and image augmentation strategies are applied to enrich the training procedure with more sufficient polyp's samples. The experimental results show that the proposed CNN model can achieve an accuracy of 98.4%, which emphasize its efficiency for helping endoscopic physicians to classify polyps and decrease the colorectal polyp miss rate.

Keywords: Colorectal cancer, Deep convolutional neural networks, DeepCPC, Polyps.

DeepCPC: نموذج التعلم العميق لتصنيف سلائل أورام القولون والمستقيم

إعداد : ديمة حسين طه

إشراف : الدكتور أحمد غازي الزعبي

الملخص

سرطان القولون والمستقيم هو مرض صامت يهاجم دون سابق إنذار، ولكن في كثير من الحالات، يكون العلاج ممكنًا عند اكتشافه مبكرًا. تلعب اختبارات الفحص دورًا مهمًا في تحديد الاورام الحميدة قبل أن تصبح سرطانية، حيث يعتبر تنظير القولون أكثر فعالية مقارنة بالاختبارات الأخرى. على مدى العقود القليلة الماضية، كان تنظير القولون والمستقيم بمساعدة الكمبيوتر موضوعًا للبحث، وحقق تقدمًا كبيرًا. ومع ذلك، لا يزال التصنيف التلقائي للأورام في الوقت الفعلي يمثل تحدي صعب، بسبب استخدام الأساليب اليدوية التي لا توفر خصائص تمييزية للصورة.

أظهرت الشبكات العصبية التلافيفية العميقة (DCNN) ثورة مهمة أثرت بشكل إيجابي على العديد من المجالات بما في ذلك تصنيف الصور. في مجال صور تنظير القولون، يمكن أن تؤثر العديد من القيود على تصنيف الاورام المستندة إلى DCNN خاصةً عدم وجود كمية كافية من عينات التدريب للمرضى، وعدم كفاية وقت التدريب، والموارد اللازمة للشبكات العصبية.

في هذه الرسالة، نهدف إلى تطوير نموذج التعلم العميق لتصنيف الاورام الحميدة في القولون والمستقيم على أساس الخصائص التمييزية المستخرجة من الشبكات العصبية التلافيفية العميقة والذي سيشار اليه باسم DeepCPC .

على وجه التحديد، نستخدم بعض نماذج الشبكات العصبية التلافيفية التي تم تدريبها مسبقًا على الصور ذات الأغراض العامة، لتطبيق مخطط نقل التعلم في نظام تصنيف الاورام الحميدة. تم تحقيق ذلك من خلال ربط مجموعة من خصائص الصور التمييزية المستخرجة من تفعيل الطبقات التلافيفية، ثم ضبط بنية DeepCPC على صور الاورام من خلال إجراء تدريب كامل من طرف إلى آخر. يتكون نموذج DeepCPC من ثلاثة مكونات رئيسية: الطبقات المنخفضة المضبوطة بدقة، وناقل الصور المسلسل، والطبقات العلوية المتصلة بالكامل.

تُستخدم مجموعة بيانات CVC-ClinicDB لتقييم نموذج DeepCPC ، ولكن يجب تطبيق المزيد من استراتيجيات استخراج التصحيح وزيادة الصورة لإثراء التدريب بعينات أكثر من الورم. أظهرت النتائج التجريبية ان DeepCPC يمكن أن يحقق دقة تبلغ 98.4٪، مما يؤكد على كفاءته في مساعدة الأطباء بالمنظار لتصنيف الاورام الحميدة وتقليل معدل فقد ورم القولون والمستقيم.

الكلمات المفتاحيه : سرطان القولون والمستقيم, الشبكات العصبية التلافيفية العميقة ,الأورام الحميدة.

Table of Contents

Authorization ii
Examination Committee Decision iii
Dedication iv
Acknowledgement v
Prepared by vi
Abstract vi
Abstract in Arabic viii
Table of Contents x
List of Figures xiii
List of Tables xiv
Table of Abbreviations xv
Chapter 1 Introduction 1
1.1 Overview 2
1.2 Problem Statement 5
1.3 Research Questions 5
1.4 Research Objectives 6
1.5 Motivation 6
1.6 Research Scope 6
1.7 Research Contributions 7
1.8 Thesis Outline 7
Chapter 2 Background and Literature Review 9
2.1 Clinical Background 10
2.1.1 Colorectal polyp 10
2.1.2 Colorectal Cancer Screening and Surveillance 11
2.2 Machine Learning Background 13
2.2.1 Supervised Machine Learning 13
2.2.2 Unsupervised Machine Learning 14
2.2.3 Reinforcement Machine learning 14
2.3 Artificial Neural Networks 14

2.4 Convolutional Neural Networks 16
2.4.1 Convolutional Neural Network Functions 17
2.4.2 Hyperparameters 18
2.4.3 Training Alternatives 18
2.4.4 Regularization Techniques 19
2.5 Related Works for Polyp Classification 20
2.6 Summary 26
Chapter 3 Methodology and DeepCPC Model 27
3.1 Datasets 28
3.1.1 ImageNet (Deng et al., 2009) 28
3.1.2 CVC-Clinic DB (Bernal et al., 2017) 28
3.2 The DeepCPC Model 29
3.3 Image Preprocessing 31
3.4 Model Initialization 32
3.4.1 Benchmark CNN Architectures 32
3.4.2 CNNs Comparison Procedure 34
3.4.3 Top Layers Configuration 34
3.4.4 Transfer learning and Fine Tuning 35
3.5 Features Concatenation 35
3.5.1 Features Extraction and Selection 36
3.5.2 Feature Concatenation Procedure 36
3.6 Model Performance Evaluation 37
3.7 Summary 39
Chapter 4 Implementation and Results 40
4.1 Experiments Configuration 41
4.2 Programming Language and Libraries 41
4.2.1 Keras 41
4.2.2 TensorFlow 42
4.2.3 Python for Data Science 42
4.3 Input Data Preparation 42
4.1.1 Patch Extraction 42

4.1.2 Dataset Split 43
4.4 Experiments of Model Initialization 44
4.4.1 Hyper-parameters Optimization 44
4.4.2 Transfer Learning 44
4.4.3 Experimental Results of Model Initialization 45
4.5 Features Concatenation Experiments 47
4.5.1 FT-Xception Setups 47
4.5.2 FT-MobileNetV2 Setups 47
4.5.3 Hyper-parameter Initialization 48
4.5.4 Final Model Fine-tuning 48
4.5.5 Results and Discussion 48
4.6 Summary 52
Chapter 5 Conclusion and Future Work 53
5.1 Conclusions 54
5.2 Future Work 55
Bibliography 56
Appendix A: Specifications of Pretrained CNNs Architectures 64

List of Figures

FIGURE 1. 1. ENDOSCOPIC APPEARANCE OF COLORECTAL CANCER: 3
FIGURE 2.1. PARTS OF THE COLON WITH POLYP 10
FIGURE 2.2. IMAGES TAKEN BY A COLONOSCOPE 11
FIGURE 2.3.WIRELESS CAPSULE ENDOSCOPY 12
FIGURE 2.4. IMAGES TAKEN BY A WIRELESS CAPSULE. 12
FIGURE 2.5. BIOLOGICAL NEURON. 15
FIGURE 2.6. ARTIFICIAL NEURAL NETWORK. 15
FIGURE 2.7. TYPICAL DEEP CNN WITH 2 HIDDEN LAYERS, 2 POOLING LAYERS, FULLY CONNECTED LAYER, AND 9OUTPUT UNITS. 16
FIGURE 3.1. SAMPLEIMAGES OF CVC-CLINIC DATASET. ORIGINAL IMAGES SHOWN IN THE FIRST ROW ANDTHEIR CORRESPONDING GROUND TRUTH SHOWN IN THE SECOND ROW. 29
FIGURE 3.2. A DEPICTION OF THE DEEPCPC MODEL FOR COLORECTAL POLYP CLASSIFICATION. 30
FIGURE 3.3. SAMPLE AUGMENTED POLYP IMAGES WITH DIFFERENT SHAPES AND SIZES. 31
FIGURE 3.4. TRANSFER LEARNING APPLIED ON CNN ARCHITECTURE. 35
FIGURE 3.5. THE PROPOSED PROCEDURE OF FEATURES CONCATENATION. 36
FIGURE 4.1. ILLUSTRATE PATCH EXTRACTION FOR POLYP AND NON-POLYP. 43
FIGURE 4.2. ROC CURVES FOR POLYP CLASSIFICATION. 46
FIGURE 4.3.THE PERFORMANCE RESULTS USING CONCATENATION FEATURES. 49
FIGURE 4.4. TRAINING AND VALIDATION ACCURACY RESULTS. 50
FIGURE 4.5. CONFUSION MATRIX OF DEEPCPC. 51
FIGURE 4.6. CONFUSION MATRIX OF FT-XCEPTION AND FT-MOBILE NETV2. 51
FIGURE A.1. VGG-16 AND VGG-19 ARCHITECTURE. 64
FIGURE A.2. INCEPTION-V1 ARCHITECTURE. 65
FIGURE A.3. INCEPTION-V3 ARCHITECTURE. 66
FIGURE A.4. RESNET ARCHITECTURE. 67
FIGURE A.5. XCEPTION ARCHITECTURE. 68
Figure A.6. MobileNetV2 blocks Architecture 69

List of Tables

TABLE 2.1. SUMMARY OF APPROACHES PROPOSED FOR COLORECTAL CANCER CLASSIFICATION. 24
TABLE 3.1. A SUMMARY OF CNN MODELS CHARACTERISTICS 34
TABLE 3.2. THE CALCULATION OF TP, TN, FN, AND FP 38
TABLE 4.1. TRAINING-TESTING SPLIT DETAILS. 44
TABLE 4.2. HYPER PARAMETERS USED IN MODEL INITIALIZATION. 44
TABLE 4.3. EVALUATION OF CNN'S MODELS USING CVC-CLINIC DB IMAGES AS INPUT. 45
TABLE 4.4. PERFORMANCE COMPARISONS OF THE PROPOSED MODELS. 52

Table of Abbreviations

Abbreviations	Full Form
AI	**A**rtificial **I**ntelligent
CAD	**C**omputer **A**id **D**iagnosis
CNN	**C**onvolutional **N**eural **N**etwork
CRC	**C**olorectal **C**ancer
CT	**C**omputed **T**omography
DL	**D**eep **L**earning
DNA	**D**eoxyribo**n**ucleic **A**cid
DCNN	**D**eep **C**onvolutional **N**eural **N**etwork
FCN	**F**ully **C**onvolutional **N**etworks
FC	**F**ully **C**onnected
GI tract	**G**astro**i**ntestinal **T**ract
ILSVRC	**I**mageNet **L**arge **S**cale **V**isual **R**ecognition **C**hallenges
ML	**M**achine **L**earning
SSA	**S**essile **S**errated **A**denoma
SVM	**S**upport **V**ector **M**achine
ReLU	**R**ectified **L**inear **U**nit
VGG	**V**isual **G**eometry **G**roup
WCE	**W**ireless **C**apsule **E**ndoscopy

Chapter 1

Introduction

This chapter presents the importance of research involved in the thesis and provides a general introduction about the challenges which aim to tackle in the domain of polyps image classification using deep neural networks. Additionally, the problem statement is formulated and discussed based on some defined objectives, questions, and limitations that triggered the motivation to address them in this work's contributions. The remaining of this chapter is organized as follows: Section 1.2 introduces the problem statement of this research; Section 1.3 presents the research questions; Section 1.4 summarizes the study aims; Section 1.5 presents the study motivation; Section 1.6 provides the study scope and limitation; Section 1.7 discusses the study contribution and its significance; Finally, Section 1.8 presents the thesis outline.

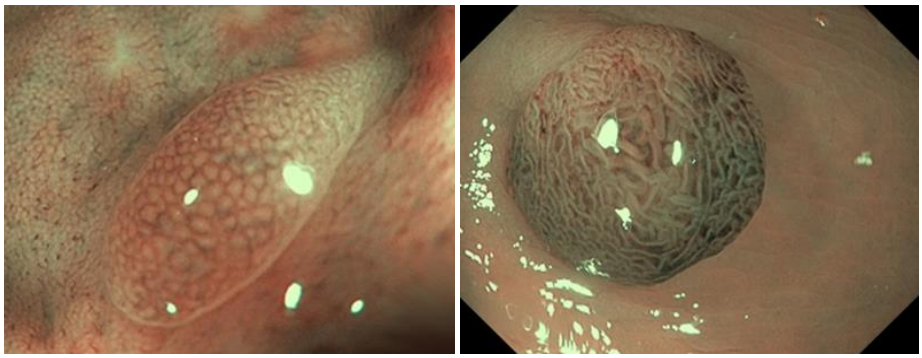
1.1 Overview

Among different types of diseases associated with gastrointestinal, ranges from annoyances to lethal diseases, colorectal cancer is the second most common cause of cancer-related death in the world according to statistics provided by the international agency for research on cancer (Ferlay et al., 2019), which shows over 1.8 million new cases in 2018.

Colorectal adenomas polyps are small neoplasm of cells formed in the lining of the colon (ACS, 2018). In spite of the fact that most colon polyps are harmless and do not usually cause symptoms; some may become cancerous over time (Groff, 2008). The process of transforming colorectal adenomas to cancerous adenocarcinoma is slow. Therefore, the early detection, endoscopic and histopathological characterization is crucial in preventing colorectal cancer and treating them safely (Dekker & Rex, 2018).

Colorectal cancer screening is a key principle in detecting polyps and avoiding the development process (Arnold et al., 2016). There are several popular screening tests for colorectal cancer, such as stool DNA tests, computed tomography colonography, wireless capsule endoscopy, and colonoscopy. Currently, the most commonly used screening test for colorectal polyps is colonoscopy (Lieberman et al., 2012). However, the challenging part of such test is the follow-up procedure after screening or CRC surveillance, where no polyp is detected but some undetected polyps may develop to another stage before conducting the next checkup.

Several studies inclouding (Rex et al., 1997; Huang et al., 2012) have shown that colonoscopists usually miss small colorectal polyps and even some larger polyps. Moreover, producing an accurate characterization of polyp's type is a challenging task, e.g. recognizing the difference between sessile serrated polyp and hyperplastic, shown in Figure 1.1, which is not well suited in any screening method (Kahi, 2015).



A) Hyperplastic polyp B) Sessile Serrated Adenoma

Figure 1.1. Endoscopic appearance of colorectal cancer:
(A) Hyperplastic polyp (B) Sessile Serrated Adenoma[1]

The decision of removing or ignoring polyps imposes expertise beyond the proficiency of many endoscopists, thereby becoming a challenging task for pathologists (Irshad et al., 2014). The reason is that the biopsy requires the patients undergo a colonoscopy examination while cutting off the small samples of suspicious parts of the polyp after detection, which should be examined later by the histopathologic. This process is obviously inconvenience and time-consuming in classifying the polyp type by practitioners (Horv´et al., 2016).

In response to the early diagnoses and prevention of colon cancer, urgent actions are needed by physicians and computer vision researchers in order to tackle the limitations and to find more accurate and reliable detection and classification methods of polyps based on endoscopic images.

[1] Images from endoscopy campus website by https://www.endoscopy-campus.com/en/classifications/polyp-classification-nice/

ML has been tremendously developed as one of the essential solutions utilized in various intelligent information technologies (Srivastava et al., 2011). Machine learning algorithms have been constrained in several aspects, e.g. speed and structure complexity. However, the recent advances in the capabilities of machines and processors have enabled developing computer-aided diagnostic systems with high performance, which also have served diagnostic processes in many health-related realms.

Various CAD based on machine learning approaches for tumor detection have been actively investigated in colonoscopy (Jerebko et al., 2003; Summers et al., 2005), but few of them have been evaluated in clinical settings (Mori et al., 2017). Basically, the CAD system extracts the features from polyp images using the hand-made features such as color, texture-based, and local binary pattern (Ameling et al., 2009; Karkanis et al., 2003). These features can be used for classification, e.g. classifying abnormal tissues into lesion or non-lesion, malignant or benign, which have been obtained from segmented objects using particular input features, e.g. area, contract.

However, the automation of polyps recognition is a technical challenge in practice since the same type of polyps can vary significantly in size, color, and texture (Pogorelov et al., 2019). Additionally, many polyp types do not clearly emerge from the surrounding mucosa; therefore, such automated methods may cause performance degradation due to the similarity of feature pattern in polyp and non-polyp parts.

Deep learning elevated to the prominent position in the field of endoscopy, when neural networks started outperforming other traditional methods on a wide range of screening tests (Zhou et al., 2017). For instance, a noticable success has been achieved in recognizing anatomical locations in diagnosis, polyp image analysis benchmarking, and polyp's detection and classification. Various deep learning methods based on convolutional neural networks (CNN) suggest a new insight into colorectal cancer diagnoses including endoscopy analysis tasks to overcome their limitations and improve the performance in several applications.

The majority of existing studies are mainly focused on polyps detection, but few of them were focused on polyps classification. This work proposes a CNN-based deep learning model to generate and learn discriminative image features based on concatenating image features for automated classification of colorectal polyps in colonoscopy images by using different configuration and distinct CNN network.

1.2 Problem Statement

In general, the major problem in the colorectal polyp diagnosis process is the dependency on the specialist's experience to primarily decide the final diagnosis (Horv´ath et al., 2016). Handcrafted analysis methods can be expensive, tedious, and error prone. Consequently, the evolution of intelligent image-based diagnostics has gained a remarkable attention by researchers and medical practitioners.

The availability of devolving deep learning CNN models and powerful machines has accelerated the progression of new intelligent diagnoses. These diagnoses are dedicated to analyze images and implement an accurate classification of any potential lesion areas, which provides a clinical decision about patients' health. Nonetheless, deep CNN can be seriously over fitted by biomedical databases that typically have hundreds or thousands of images.

Additionally, there are some challenges in developing DL algorithms and integration a variety of polyp features into one system in order to classify particular polyp types, such as the limitation of polyp images makes CNN training for automated colonoscopy image classification so difficult, also the learning from scratch can be time-consuming. Consequently, a deep neural network model with adequate feature extraction and notable feature representation needs to be utilized.

1.3 Research Questions

To attempt addressing the limitations discussed in the previous sections, the following specific questions are posed:

1. Does the mechanism of transfer learning provide satisfied classification results when the fine-tuning is applied with insufficient training dataset?
2. What is the impact of extracting deep image features from the activations of convolutional layers on the accuracy of polyps classification?
3. Does the CNN models performance be further maintained by adding optimized fully-connected layers to project image features into low-dimensional spaces?

1.4 Research Objectives

The aim of this research is to produce an accurate model that is most suitable for aiding the doctors to classify polyps and make decisions at early stages. To reach this aim, the following specific objectives are posed.

1. To provide an automated CNN-based model for colorectal polyps classification diagnosis.
2. To explore the transfer learning approach on various CNNs model with polyp dataset and fine tuning.
3. To improve the extracted image discriminative features extremely that increases the model performance.

1.5 Motivation

Recent trends in research expounded that deep learning networks are very influential for automatic image analysis. Deep learning models have shown great improvements in the performance of numerous medical applications. This possess our curiosity to find and perform an effective method based on DL algorithms to determine the discriminative features in colorectal polyp images, which in turn will lead to an effective classification model that can potentially be a life savior.

1.6 Research Scope

The scope of this thesis is to provide insights on the CNN-based deep learning architecture and features appropriate for polyp classification tasks. It focuses on the analysis of features extraction and concatenation from activations of convolutional layers using transfer learning. Several pretrained deep CNN architectures utilized in these experiments including: VGG19, InceptionV3, ResNet50, Xception, MobileNetV2 and ResNet152V2. Moreover, the model training and testing was performed on the public CVC-Clinic dataset (Bernal et al, 2017).

1.7 Research Contributions

The main contributions of this thesis can be summarized in the following:

- A deep CNN-based model is proposed to generate discriminating image features that are extracted from different predefined architectures then concatenated into a single image descriptor.
- An in-depth performance analysis is provided using different CNN architectures and setups in order to demonstrate the effectiveness of utilizing the transfer learning from source domain (general images) to target domain (polyp images) for existing CNN hyperparameters in tackling the lack of training data.
- Empirically, this research has shown the efficiency of the proposed learning model in the context of selected features, layers, and hyperparameters on the classification polyp images, which may inspire more research works to investigate its performance in the domain of medical imaging where sufficient data can hardly be obtained.

1.8 Thesis Outline

Chapter 1 provided a general introduction for the domain of polyps' image classification using deep neural networks. The research problem, objectives, limitations, and scope are also discussed. The rest of this thesis is organized as follows:

Chapter 2 introduces a clinical background of colorectal polyps and screening tests. Also, it presents comprehensive details on the fundamentals of machine learning and neural networks but focused on the CNN models. Numerous studies and recent related works are discussed thoroughly.

Chapter 3 discusses the proposed methodology and illustrates the proposed architecture of deep CNN-based model for polyps classification (DeepCPC). It also provides thorough illustrations about several modified CNNs utilized in this work. Additionally, it presents the dataset improved and employed to train and evaluate the performance of DeepCPC.

Chapter4 presents the environment requirement and thorough analysis on the implementation of updated CNN networks and the results. It also discusses the experimental results of the DeepCVC model with performance comparisons.

Chapter5 concludes the thesis by summarizing the findings and how they relate to the research problem and objectives. It also outlines the possible directions that could expand this research work in the future.

Chapter 2

Background and Literature Review

This chapter covers a medical background and general aspects of machine learning and convolutional neural network. Section 2.1 presents an overview of the medical scenario in the colorectal polyp with reviews of various screening methods used in the diagnosis; Section 2.2 provides a background of machine learning with a description of learning types; Section 2.3 discusses an overview of the artificial neural network mechanism; Section 2.4 explores the convolutional neural network and its primary parameters; Section 2.5 illustrates the training methods utilized in deep learning; Section 2.6 provides the techniques used to perform regularization; Section 2.7 shows previous studies associated with the ML algorithms used for colorectal polyps problem, discussing their finding and limitations; Section 2.8 provides a summary of this chapter.

2.1 Clinical Background

This section provides a clarification of polyp and process of growths, given different modalities for screening tests.

2.1.1 Colorectal polyp

Polyps are abnormal growths of tissue from a mucous membrane (inner lining) of the colon or rectum. As shown in Figure 2.1, polyps are commonly found in different parts of the gastrointestinal system, such as the stomach, ear, nose, urinary bladder, colon, and uterus. Some polyps are flat, it is said to be sessile while others attached to the surface by a narrow elongated stalk (ACS, 2018).

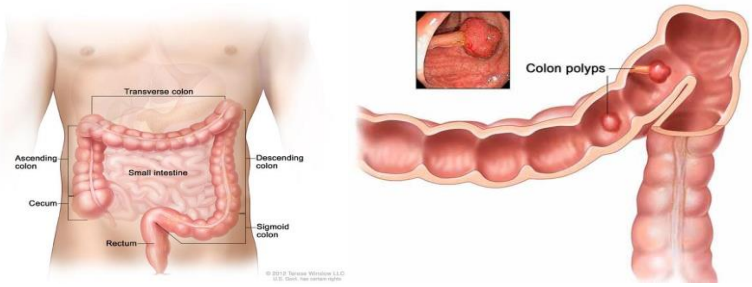


Figure 2.1. Parts of the Colon with polyp[2]

[2]Image from the National Cancer Institute © 2012 Terese Winslow LLC by https://www.cancer.gov/types/colorectal/patient/rectal-treatment-pdq

2.1.2 Colorectal Cancer Screening and Surveillance

Screening of colon cancer can detect early cancers and polyps in the large gut. This kind of screening may be able to find the problems that can be treated prior to developing or propagating cancer. Regular tests can reduce death risk and colorectal cancer complications. Several test options are available for colorectal cancer screening such as the following:

Colonoscopy is a laparoscopic examination of the lower digestive tract, including the anus, the rectum, the colon, and sometimes the last part of the small intestine (terminal ileum), during a colonoscopy, a long, flexible tube (colonoscope) is inserted into the rectum. A precise video camera on the edge of the tube filming inside the body, and connected directly to the TV screen allows the doctor to view the entire colon, If necessary, polyps or other abnormal tissue can be removed through the endoscope during the colonoscopy procedure, or to stop bleeding, as well as tissue samples (biopsies) can also be taken during colonoscopy. An image taken by colonoscopy is shown in Figure2.2.

Figure 2.2. Images taken by a colonoscope[3].

Computed Tomography Colonography a special device of radiology, gives an image of the organs of the body using x-rays taken from different angles, then with the computer assistance, these images overlap on top of each other, giving a detailed 3-dimensional image that presents any abnormalities or tumors, that cannot be shown by normal radiographs.

[3]Figure published in depeca website byhttp://www.depeca.uah.es/colonoscopy_dataset/

Wireless Capsule Endoscopy a medical procedure used to diagnose and visualize the gastrointestinal tract. The endoscopy is carried out using a small special wireless camera placed inside a capsule swallowed by the patient, shown in Figure2.3.The capsule is small in size, enabling it to reach places that the normal endoscope cannot reach, as well as the sharpness of its images and the magnification it provides.

Figure 2.3. Wireless capsule endoscopy[4].

As the capsule travels through the digestive tract, it takes thousands of pictures of the stomach, intestines and other parts, example pictures from a WCE are shown in Figure 2.4. It is recorded and sent to a receptor that is fixed around the patient's waist. The capsule stays in the patient's body for about 8 hours and goes out with the stool. It is not reused afterward.

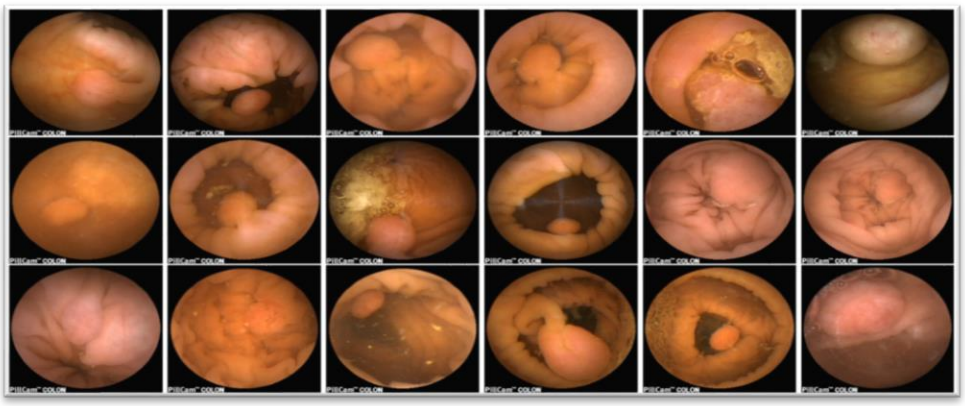


Figure 2.4. Images taken by a wireless capsule[5].

[4]Figure released into the Diagnostic and Interventional Cardiology webpage, https://www.dicardiology.com/article/capsule-endoscopy-systems-safety-patients-cardiovascular-implants

[5] Figure published in the article "Polyp Detection and Segmentation from Video Capsule Endoscopy: A Review" https://www.mdpi.com/2313-433X/3/1/1/htm.

The next following sections, will discuss the machine learning in details covering different types, then presenting the artificial neural network processes and convolutional neural network, finally illustrating the proposed methods for colorectal polyps and applications.

2.2 Machine Learning Background

Artificial intelligence has recently captured the world's attention in various forms (Shalev-Shwartz & Ben-David, 2017). Several models confirm that artificial intelligence has approached competition from human intelligence, with this success, increasingly more interest with what is known as machine learning to move towards greater gaining in employing artificial intelligence.

ML is one of AI branches based on building computer programs in various forms, which improve their performance at some task through the experience. The term machine learning was coined by the pioneer of artificial intelligence Arthur Samuel in 1959 while working in IBM laboratories (Awad& Khanna, 2015), he used the game of checkers to create the first learning program.

In the mid-90s, machine learning rose as the new substance of AI. Bigger datasets were created and made open to permit more individuals to manufacture and train AI models (Burkov, 2019). Despite the fact that ML calculations draw derivation from statistics, what make it powerful is the attempts to limit the error between the expected yields given by the dataset and predicted calculation yield to find the advanced principles. ML approaches are basically three: supervised, reinforcement, and unsupervised. These approaches are discussed briefly in the following subsections:

2.2.1 Supervised Machine Learning

Supervised machine learning algorithms perform the data mining task in which both input (training sample) and the expected output data (labels) are provided, the task of the algorithm is to analyzes and learn a mapping function from training data inputs to predict the correct label in the outputs. Supervised learning problems can be further grouped into sub-types of learning according to the required output from the machine learning system. One of the most important of these types is classification; the goal of the learning process is to produce a model that can classify any new income into one or more of the previously defined types. The other type is regression, which is similar to the classification, except that it predicts values with resides continuously instead of separate classes.

Different supervised machine learning techniques and feature selection algorithms have been widely applied to colorectal cancer by numerous studies, such as prediction of early colorectal cancer metastasis (Takamatsu et al., 2019), automated polyp detection system (Zur et al., 2018), or to clasify colorectal polyps (Komeda et al., 2019).

2.2.2 Unsupervised Machine Learning

Also called descriptive learning, unlike the previous type, unsupervised learning performed through incoming data without any predefined output, and the goal is to devise new models and hidden relationships between the data. The most important unsupervised learning method is clustering, which uses exploratory data analysis for grouping or to find hidden patterns that are not previously known. Unsupervised learning explored in colonic disease studies, e.g. clustering proteomic data from colon cancer (Zufferey etal., 2018), radiomic features of quantitative imaging phenotypes (Chen et al., 2019).

2.2.3 Reinforcement Machine learning

It is concerned with how software agents have to take actions on the current behavior. The agents learn by giving signals symbolizing reward or punishment based. Semi-supervised learning is between supervised learning and non-supervised learning, where it gives an incomplete training signal: a training group with some (often) of the missing outputs targeted. Transduction is a special case for this principle where a whole set of problem states are known as learning time, except that part of these goals is missing. Semi supervised is utilized for metastatic colon lymph node diagnosis (Michaeli et al., 2012).

2.3 Artificial Neural Networks

They are synthetic neurons similar to the neurons in human nervous system, especially in the brain as shown in Figure 2.5. What happens in human brain as a permanent activity is an electrochemical activity between networks of brain cells called neurons or nerves cells. In particular, the neuron consists of many inputs that are responsible for communication called (dendrites), and one output (the axon) that links to different neurons or biological tissues. The signals are transmitted from one neuron to others by the synaptic method. The signals may be inhibitory or exciting. If the

signal is exciting (the ability to move) then it activates the neurons and spreads via the axon. (Nicholls, J. &Kuffler, S., 2018).

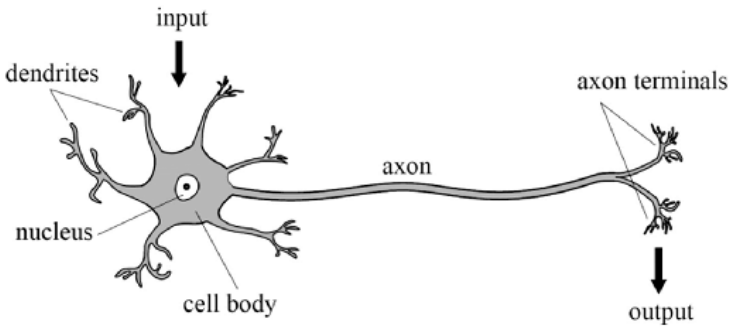


Figure 2.5. Biological neuron[6].

Accordingly, artificial intelligence has tended to build artificial neural networks that mimic the human brain. Artificial neural networks consist of neurons that communicate with each other via links. These links are used to transfer signal between these neurons, and each link has a specific weight that increases with the rise in the strength of the contact between the two neurons connected via this link.

As shown in Figure 2.6, the data that need to be process is placed at the first layer of units (x_1, x_2,…x_i), each input is connected to neurons (j) and weights (w_{1j}, w_{2j},…w_{ij}). Every signal being multiplied with its related weights on the connection, then the neuron sums all signal received and send to the non-liner activation function φ(•) in order to give the final output (y).

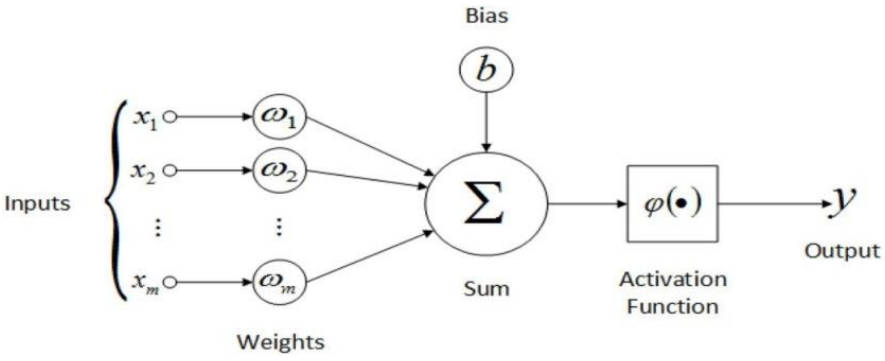


Figure 2.6. Artificial neural network[7].

[6]Neves, A. C., González, I., Leander, J., & Karoumi, R. (2017). A New Approach to Damage Detection in Bridges Using Machine Learning. Experimental Vibration Analysis for Civil Structures, 73–84.
[7]Figure published in the article on the medium.com website. https://medium.com/@jayeshbahire/the-artificial-neural-networks-handbook-part-4-d2087d1f583e

2.4 Convolutional Neural Networks

DL has become the latest model for many image and signal processing tasks. DL is a new branch of ML, which relies on a set of algorithms based on modeling high-level abstractions in data by extracting multiple processing layers, allowing systems to be able to learn complex mapping functions directly from input data, thereby achieving AI goals, which was recognized as one of the top 10 breakthroughs in 2013(Person, 2015).

Deep CNN model is a multi-layer neural network in which there is a connectivity pattern between its neurons, each neuron receives an input, a dot product between each input and its associated weight is performed, followed with a non-linearity. There are two main parts for CNN: A convolution tool that splits the various features of the image for analysis, a fully connected layer that uses the output of the convolution layer to predict the best description for the image as shown in Figure 2.7.

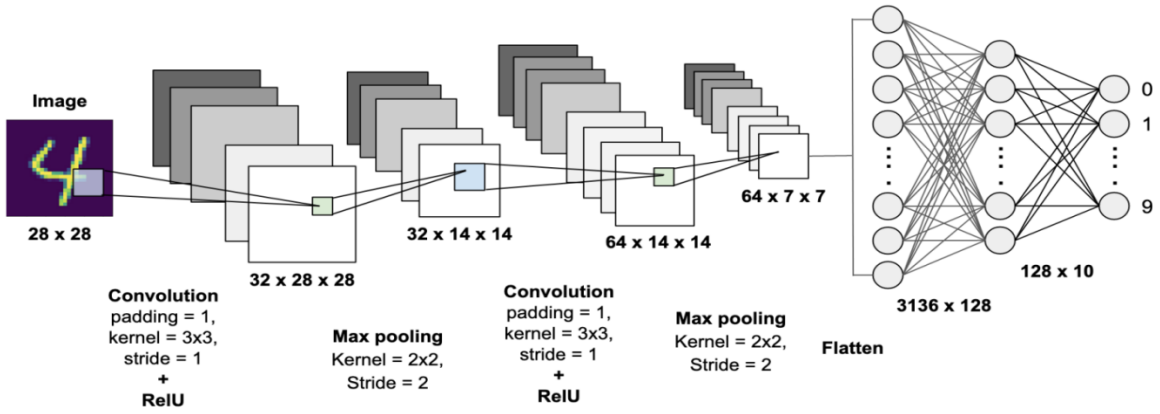


Figure 2.7. Typical deep CNN with 2 hidden layers, 2 pooling layers, fully connected layer, and 9output units.[8]

Neurons are arranged in three dimensions – width, height, and depth, within the layer also neurons are connected to a small area of the layer before them, called the receptive field, and are not as connected as in a normal neural network. The structure of CNN networks consists of several

[8] Figure published in the towardsdatascience article web page https://towardsdatascience.com/mnist-handwritten-digits-classification-using-a-convolutional-neural-network-cnn-af5fafbc35e9

different types of sequential layers, and some will be replicated. The following are some of the common CNN layers:

Input Layer is an image that has the following dimensions width x height x depth, saved as a matrix of pixel values, an RGB image which has been separated by its three color planes: Red, Green, and Blue.

Convolutional layer is the basic building block of CNN, consists of a set of filters that are grouped across the width and height dimensions of the image. Filters in which the image is converted have the same number of dimensions as the image, and usually used spatial sizes such as 3x3 or 5x5. The width and height of the output depending on the size of the filter, the stride which is the number of pixels in which the filter is transferred between each calculation, usually one or two, and the number of zero padding around an image. The output depth will be equivalent to the number of filters applied.

Pooling layer provides a method for non-linear reduction of volume using small filters for sampling, two common methods are average pooling and max pooling aggregation in a rectangular area of output from the previous layer, therefore reduces the number of parameters and calculations, spatial size, as well as avoids over fitting.

Fully connected layer takes the end results of the convolution and pooling layers process and reaches a classification decision; consist of input layer that takes the output from previous layer and turns them to single vector (flatten), that can be used as an input for the next fully connected layer that takes the input from features analysis and perform weights to predict correct label, last output layer gives the final probabilities of certain feature belongs to a label.

2.4.1 Convolutional Neural Network Functions

After introducing the basic structure of CNNs, we will present some of core functions in CNN-based models:

Activation function is a mathematical equation that squashes the input values into a certain range to determine the output of a neural network, each neuron in the network attached to this function which then determines whether to be activated or not based on the relevance between the inputs from each neuron with the model prediction. Sigmoid, tanh, and rectified linear unit (ReLU) are widely-used activation functions, as they prevent saturation problems and make learning quicker than most

functions. Mathematical utility is often employed by sigmoid (nonlinear) functions since they are very simple to compute derivative, and we use them to measure weight updates for training algorithms.

Loss function is a method for evaluating the error between the output and the given target value, it is used to determine how well the algorithm models, the data, and what needs to be optimized.

Softmax is a factor which normalizes an input into a value vector that follows a distribution of probability with a total of up to 1. The value of the output is between [0,1].

2.4.2 Hyperparameters

A parameter set prior to the learning process. These parameters can be tuned and largely affect the model performance. Among important parameters are:

Learning rate: is a parameter controls how much the model will adjust when the model weights are changed, in response to the expected error. Choosing the learning rate is difficult because a value that is too small may result in a lengthy training cycle that may get stuck, while a value that is too high might result in learning a sub-optimal set of weights that are too fast or unstable.

Batch size: refers to the number of samples examples utilized in the one iteration belong to such subset.

Epochs: indicates the number of passes of all the data, meaning that one pass is a full learning cycle of entire training dataset prediction, along with weight update and cost calculation.

2.4.3 Training Alternatives

There are three common approaches utilized in deep learning to perform classification tasks:

1- **Training from scratch**: Requires a very large labeled data set to be gathered, in addition, network architecture needs to be designed to learn features and models. This is perfect for new apps or apps with a large number of output categories. This is a less popular approach since, with a large amount of data and a high learning rate; such networks typically take days or weeks to train.

2- **Transfer learning:** Recently, it observed that an increasingly applied to deep neural networks in many domains, thus to develop a deep learning network with an acceptant result, requires large amounts of training data, and significant training time (Sarkar et al, 2018). Transfer learning is the motivated method to solve one problem from the pre-trained model. It gains the parameters and weights from an already-trained network and used as the starting point on a new problem, therefore modify them to be fit in the target model. Transfer learning may also assist the target project learning within the following ways:
   - Stepped forward baseline performance: when we increase the understanding of an ignorant learner with the realization from a source model, the baseline overall performance may enhance due to this expertise transfer.
   - Development time of the model: adapting information from a source model may additionally assist in completely gaining knowledge of the target task, compared to learn from scratch. Therefore, improve the total time taken to develop or learn a model.
3- **Fine-tuning:** A process to train the pre-trained network on a small dataset. Usually, fully connected layers, which can be used as classification layers, the pre-trained layer will obtain a lower learning rate. This will enable the features to be adapted to the new dataset.

2.4.4 Regularization Techniques

One of the most common problems facing data science professionals is to prevent over-fitting. When the model performed extremely well on train data, but was unable to predict test results. Different regularization techniques to overcome the overfitting issue such as:

Data Augmentation: this is a technique that allows increasing the data available for training models without collecting new data. This can be in different schemes relying on the dataset used. Techniques for data augmentation, such as cropping, padding, and horizontal flipping are widely used to train large neural networks.

Dropout: the main concept is to deactivate input units by random after each iteration. The set of 2^n neural networks is a neural network with n units that hires dropout. Such alternative networks have a reduced number of units but also exchange weights in order to retain an equivalent overall number of parameters. Training a weight loss network can be seen as training a series of 2^n smaller weight share networks.

Early Stopping: during training, the model is tested on dataset after each epoch. When the output of the model on the validation dataset begins to degrade such as the accuracy begins to decrease or the loss begins to increase, therefore, the training phase is stopped.

2.5 Related Works for Polyp Classification

This section, discuss the related work in regards to colorectal polyp classification, the methods employed, introduce the experiments were performed, and the results.

A pipeline for image classification can be classified into extraction features and classification training. Conventional feature extractors are mostly based on pre-designed techniques, which require a strong knowledge of the field and considerable technical skills. In addition, the workload is heavy when multiple parameters are finely tuned. The previous study on the diagnosis of colorectal polyp focused on techniques of feature engineering, for example, to extract features from vessel structures (Tamaki et al., 2013), or using scale-invariant key points sampled from local patches as characteristics such a study proposed by (Stehle et al.,2009),they perform classification on colon polyps as a multi-stage system, by compared two learning algorithms in order to perform vessel segmentation, after that used the results as seed for fast marching algorithm to carry out other segmentation for the whole vessel lumen. Consequently, the computed features from segmentation used to classify the polyps. The system achieved a correct classification rate of 90% evaluated on the datasets that include 56 polyps with histologically confirmed ground truth. A limitation of these studies is that each region of interest had to be selected manually.

Other study focused on machine learning methods based on SVM, utilizing shape-based approach to extract characterizing features (Li et al., 2004). The study proposed supervised learning methods to detect the abnormal regions in colonoscope, the method extracts different size patch from the image with the resolution of 256 × 256 pixels, then SVM classifiers are trained for each size independently, the features are passed to ensemble classifier to perform classification score, then aggregate scores for a final decision. The classification task compared three methods SVM, Gaussian kernel, and MFNN, to perform classification between abnormal and normal regions. Study results show that SVM achieved a higher classification accuracy rate than others.

Cao et al., (2009) present an automated detection of the shape of the opening appendix in a colonoscopy video frame. The study proposes several intermediate-level features suitable and uses a K-mean classifier to finally classifying images into two categories: appendix image and non-appendix image, a test images consist of 800 images taken from 5 colonoscopy videos. The technique has an average accuracy of 90% in both classes.

Manivannan et al., (2013) proposed two methods for representing intermediate-scale features to classify normal-abnormal in colonoscopy images, using SVM for features selection patch-based method, and scale-space method are proposed, by using cross-validation for the patch-based method found that window size 60 perform the best accuracy and for the scale-space method, 3-level Gaussian scale-space gave the best accuracy.

Other study investigates the advanced precancerous colon lesions proposed by (Hilsden et al., 2018). The process of using ColonFlag algorithm that combines patient data to be classified based on the majority lesion found into five groups, invasive colorectal cancer, high-risk precancerous lesions, non-advanced adenomatous polyp or non-dysplastic sessile serrated polyp, non-neoplastic findings and, the last group if there's not any finding. Based on the data of the regularly collected CBC and data such as patient age, gender, and colonoscopy, the ColonFlag model was able to classify individuals at risk of a CRC. As the authors indicate, ColonFlag can also identify screened people at increased risk to the CRC, and individuals can, therefore, be targeted for greater compliance with conventional screening tests.

Recent applications with DL methods have mostly overcome the difficulties of designing these modalities, highlighted the significance success in classifying massive amount of objects in colorectal domains. Different CNN studies have been proposed for classification colorectal dieses. (Ribeiro et al., 2016) explored the classification of colonic polyps by several CNNs to classify the colonic mucosa into healthy and abnormal. The dataset consists of 25 healthy images from 18 patients and 75 abnormal images from 56 patients. The CNN’s used small patches in order to increase the dataset and to classify different regions in the same image, thus 800 images resulted to train the classifier. The results show experimentally the overall accuracy was 90.96% while the sensitivity was 95.16%, but the false positive rate was high resulting in specificity only 74.19 %.

Another used of benchmark methods such a study presented by (Byrne et al., 2017). The study used inception network architecture with stochastic gradient descent to classify each input frame into

one of four different categories. The study applied a mini-batch of 128 frames and data augmentation procedure on the dataset. Image frames for narrowband image video and unaltered routine video used in model training and validation. The CNN network results in the study can identify polyps with an accuracy of 94% percent, a sensitivity of 98%, and specificity of 83%, all with 95% CI (confidence interval).

Tian et al., (2019) proposed one-stage detection and classification approach for the five class polyp classification. The CNN network used Retinanet50 for detection and classification in one stage and two stages with a dataset containing 871 images of colorectal polyps annotated by a professional medical practitioner. In addition, the study utilized a data augmentation during training. The study measures the results with state-of-art based on the MICCAI 2015 polyp challenge. The results show that the one-stage approach is more efficient than the two-stage.

The CNN method of colon polyp CAD is used for another analysis by (Komeda et al., 2019) a total of 1,200 colonoscopy images in the datasets used in the analysis, were collected from videos of actual endoscopic examinations. The CNN network consists of several conventional layers, pooling layers, and Softmax classifier. The input image is 256 × 256 pixels, performing 10-hold cross-validation in the training process. The study reveals that the CNN-CAD method can precisely differentiate between adenoma and non-adenoma polyps up to 70%.

The research by (Wang et al., 2019) uses the CNN system intended by an automated polyp detection system to investigate the impact of ADR. 1.130 patients involved in the study with eligibility requirements. At the end of the test, the CAD colonoscopy was shown to raise ADR by 20.3% ($p<0.001$) significantly to 29.1%, relative to normal colonscopic ADR. 185 adenomas were detected and 114 polyps were detected in CAD colonoscopy. Such results are substantially higher than those found by Standard Colonoscopy relative to 102 adenomas and 52 hypetlastic polyps.

An alternative way is to adopt a transferlearning strategyfor classification the celiac diseases on the basis of the endoscopic dataset explored by (Wimmer et al., 2016). The transfer learning from CNN demonstrated great potential for the classification of celiac diseases on the basis of the endoscopic dataset. The endoscopic images have been identified in three different transfer learning techniques. CNN with fine-tuning obtained the highest classification accuracies, even though the small number of available training dataset resulted in overfitting. These were the highest results for

the diagnosis of the disease achieved in the VGG16 network with an accuracy of 90.5% in comparison to four state-of-the-art networks.

The use of transfer learning to increase adenoma detection rates was research by (Urban et al., 2018). Various CNN models performed in the study included pre-trained models for the image recognition and untrained models. The datasets selected included 8,641 colonoscopies images from 2,000 patients, the second dataset contains 1,330 colonoscopy images from a variety of patients, and another dataset includes 9 colonoscope videos. The study performed cross-validation by training the model on a single dataset and checking it on a whole new data set. The findings show models previously equipped with a polyp and random images capable of detecting polyps with 96.4% accuracy and sensitivity of 96.9%. The model also predicates all polyps discovered during the analysis by experts and polyps. The studies are summarized inTable2.2.

Table2.1. Summary of approaches proposed for colorectal cancer classification.

Reference	Study Goal	Method Used	Main Findings	Limitation
Wang et al., (2019)	Detection and classification polyp and adenoma	CNN based on SegNet	CAD system increase of ADR by 50%, from 20% to 30% compare with standard colonoscopy with sensitivity 94.38%	Difficult to assess the exact contribution of the system. Lack of external validity.
Tian et al., (2019)	Detection and classification polyp	Adapting RetinaNet	Model with one-stage approach is more efficient than the two-stage. Shows smaller training needs and inference times	High rate in miss-detected polyps compare with manual detection methods
Hilsden et al., (2018)	Predicate the appearance of polyps at colonoscopy	Machine learning utilizing ColonFlag algorithm	The model identifies the patients with high rate of CRC;achieve 95% in the specificity.	Lack of patient history information that leads to high-risk polyps in the earlier years will misclassify as non-high risk
Urban et al., (2018)	Localizes and identifies polyps	Using a pre-trained CNN for detected the presence of polyps in a frame.	Identified polyps under the ROC curve of 99%, the accuracy 96.4%, false positive 7% rate.	Require large training data for pre-training.
Byrn et al., (2017)	Differentiation of colorectal polyps	Apply DCNN (Inception Network)	Model works with unprocessed frames and can operate in quasi real-time. Achieve high accuracy in sorting diminutive colorectal polyps into adenomas or polyps.	Low confidence determination. (15%) of consecutive diminutive polyps in the test set were excluded by the AI model.
Komeda et al., (2017)	Classification adenomas or polyps	Applying CNN network	The model decision correct in every 7 out of 10 cases.	Unsatisfactory accuracy for CNN-CAD system

Reference	Study Goal	Method Used	Main Findings	Limitation
Zhang et al., (2017)	Detect and classify of colorectal polyps by using low-level CNN features	Two different CNN models for detection and classification	High detection performance	Low classification performance and limited data set
Wimmer et al., (2016)	Classification of Celiac Disease	CNN and SVM	Fine-tuning the CNNs clearly achieves the highest classification accuracies	Network overfitted
Ribeiro et al., (2016)	Classification polyp into healthy and abnormal	Applying CNN network	The model performs in the accuracy 90% and sensitivity 95.16%.	High false positive rate
Manivannan et al., (2013)	Classification Colonoscopy image to normal-abnormal	SVM	Considerable accuracy improvement regardless of the features used	Limited dataset
Cao et al., (2009)	Automatic classification in colonoscopy video	Utilizing machine learning and k-mean classifier	Technique has an average accuracy of 90% for appendix images and 90% for non-appendix images.	Falsely classified several type to another due to not applying preprocessing phase
Stehle et al., (2009)	Classification of polyps	ML	The system achieved a correct classification rate of 90%.	Limited dataset and low specificity
Li et al., (2004)	Detecting and classification abnormal regions in colonoscope images.	Utilizing SVM for detection and classification	Achieved accuracy 76.3% in classification and the average detection rate is 83.3%	Numbers of cases used were limited

2.6 Summary

This chapter commenced with a brief discussion of a medical scenario for the colorectal polyps problem, which provide general characteristics of the disease and the methods of examination followed and their challenges. As well, it highlighted the concept of machine learning, deep learning, and explored the existing synergy between the fields of colorectal polyps and machine learning, by illustrating the needs for deep learning techniques to overcome a polyp-image classification limitation. Finally, extensive research in the previous studies enraptured, however, some of the published methods invariably show two common limitations: first, low classification performance, which makes these approaches unsuitable for correctly classifying the polyps; second, all studies used their own image data set which prohibits quantitative comparison. Consequently, this guides the way towards performing the DeepCPC model for colorectal polyps classification. The next chapter will introduce the methodology of this work and presents the details of DeepCPC architecture.

.

Chapter 3

Methodology and DeepCPC Model

This chapter presents the methodology used to develop the DeepCPC model for polyps classification. Section 3.1 covers a description of the dataset utilized for polyps classification model; Section 3.2 illustrates the model and main stages of the DeepCPC model; Section 3.3 provides the pre-processing as the first stage; Section 3.4 discusses the model initialization steps by comparing six pre-trained CNN models then choosing the best of them to be trained on modified dataset; Section 3.5 presents the third stage in the DeepCPC including features extraction, selection, and concatenation; Section 3.6 illustrates the performance evaluation measures; Section 3.7 summarizes this chapter.

3.1 Datasets

This section elaborates the available datasets considered in this thesis.

3.1.1 ImageNet (Deng et al., 2009)

ImageNet is a popular dataset used for different general purposes. The ImageNet Large Scale Visual Recognition Challenges (ILSVRC) annual competition assesses large-scale algorithms and neural network structure for image detection and classification. The ImageNet dataset consists of 1000 hierarchical classes, divided into 1.2 million training sets, 50,000 testing sets, and 100,000 evaluation sets. In this work, this dataset is already utilized to train the CNNs used to initialize the DeepCPC model.

3.1.2 CVC-Clinic DB (Bernal et al., 2017)

The dataset designed as the training set for MICCAI2015 and ISBI2015 sub-challenges for polyp detection in endoscopic videos. The dataset frames extracted from 29 endoscopic videos by courtesy of Hospital Clinic, Barcelona, Spain. The dataset contains 612 frames standard definition still images of 384 × 288 resolution with 31 several examples of polyps from 31 several sequences. Each frame in this dataset consists of 612 frames ground truth corresponding to a region covered by the polyp in the image, as observed in Figure 3.1.

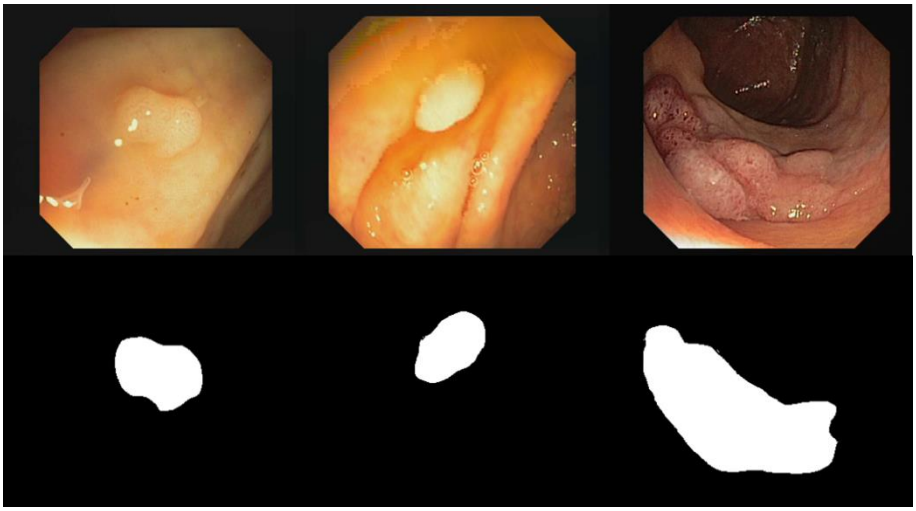


Figure 3.1. Sample images of CVC-Clinic dataset. Original images shown in the first row and their corresponding ground truth shown in the second row.

3.2 The DeepCPC Model

This work, propose a deep learning CNN-based model for colorectal polyps classification, as shown in Figure 3.2. The complete model consists of three stages: the dataset preprocessing, model initialization using pre-trained CNNs, finally, the features selection and concatenation from the activations of two well-performing model, which fed to the fully-connected layers for classification. The model details are explained in the next sub-section.

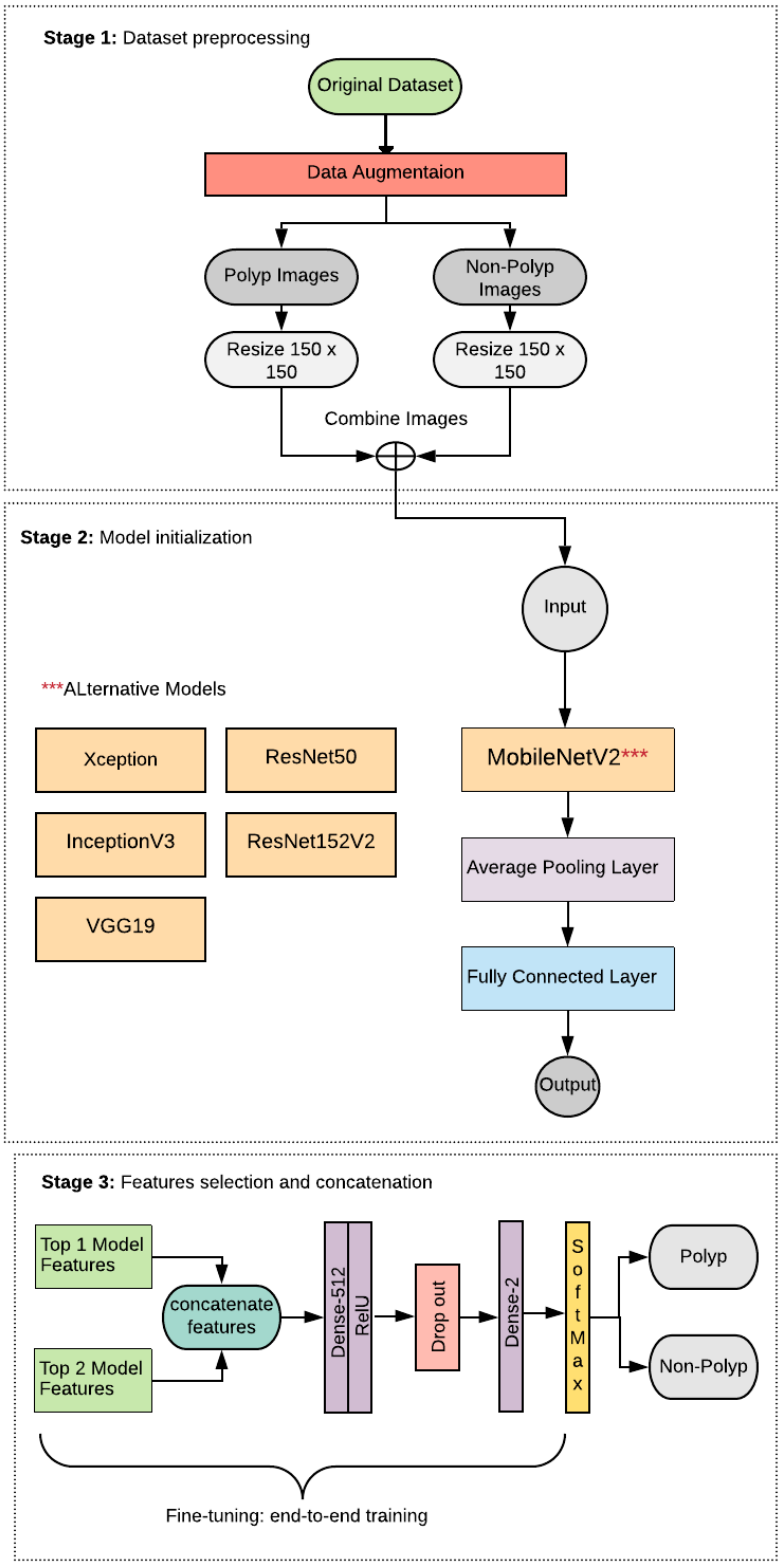


Figure 3.2. A depiction of the DeepCPC model for colorectal polyp classification.

3.3 Image Preprocessing

The available amount of training samples from CVC-Clinic dataset (Bernal et al, 2017) is insufficient. In order to perform a larger balanced dataset, therefore, a batch extraction is applied with data augmentation techniques such as rotation, flipping, scaling, and cropping. Figure 3.3 shows some sample images. Then, images are split into training, validation, and testing subsets.

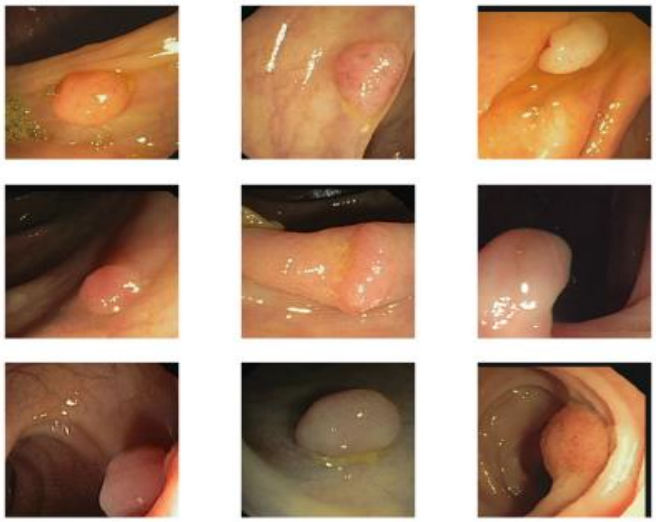


Figure 3.3. Sample augmented polyp images with different shapes and sizes.[9]

[9] Figure published in deephealthresearch website https://site.uit.no/deephealthresearch/2017/07/19/polyp-detection-using-deep-learning/

3.4 Model Initialization

The lower part of DeepCPC architecture is built from a pretrained CNN network and have the same parameters initializations of best two models that would be further processed to extract image features from them then get concatenated

3.4.1 Benchmark CNN Architectures

This section, examines some of the most popular CNN architectures used for image classification tasks: VGGNet, GoogleNet, ResNet, MobileNet, and Xception. Such networks are also used as feature extractors due to their great extraction capabilities.

VGGNet: Introduced by (Simonyan& Zisserman, 2014), consistent over multiple stacked layers, easy-to-implement architecture, and has a lot of variants based on the number of stacked layers. The VGG16 and VGG19, illustrate in Figure A.1, Appendix A. The VGG architecture comes with the 3 x 3 convolution layers, max-pooling for reducing volume size performed over a 2 x 2-pixel window with a stride of 2, in each of the hidden layers ReLU activation is used. The end of the network comes with two fully connected layers that go along with a softmax layer for prediction.

GoogleNet (Inception): Introduced by (Szegedy et al, 2014). Inspired by LeNet but implemented a novel element, the main contribution is the creation of the Inception-V1 module, which notably decreased the wide variety of parameters with the aid of forty million on the top of the convolutional layers. The architecture consisted of a 22 deep layers as shown in Figure A.2, Appendix A, which eliminated a massive variety of parameters through the usage of average pooling instead of fully connected layers. In addition, several versions of the GoogleNet were launched, such as InceptionV3 that makes use of batch normalization in particular, comes with 24M parameters, and 42 layers as shown in Figure A.3, Appendix A. The motivation of Inceptionv3 is to prevent representational bottlenecks i.e. to seriously limit the input dimensions of the subsequent layer and to have extra powerful computations by using of factorization techniques.

ResNet: Developed by researchers at Microsoft (He et al., 2015), which received first place in the ImageNet competition in 2015. ResNet architecture consists of 152 layers, includes batch

standardization and unique skip connections to educate deeper architectures as shown in Figure A.4, Appendix A. In addition, ResNet with 1000 layers can be equipped. However, it has been found that ResNet typically operates on blocks of relatively low depth 20-30 layers operating in parallel rather than serially flowing across the entire length of the network. Various ResNet models can be implemented such as ResNet-50,152, which the numbers follow the model name indicate the numbers of the layers in the architecture of the model.The ResNet-50 considered as a smaller version of ResNet 152.

Xception: Developed in Google, inspired by previously published GoogleNet's Inception architecture, the architecture is novelty the adaptation of modified depth wise separable convolution layers, which proven to outperform the InceptionV3 on ImageNet data set and achieved .94% accuracy. The model made up of 58 layers with 36 convolutional layers structured into14 blocks as feature extraction base, the middle flow of Xception network, as shown in Figure A.5, Appendix A, repeated 8 times. The blocks are separated by residual layers with total of 13 depth-wise layers' employ kernels of size 3 x 3, 14 ReLU activation layers,19 batch normalization layers,4 Max-pooling layers use a 3 x 3 kernel with a stride of 2, one global pooling layer, and one fully connected layer. The normal convolutional layers utilized both a 1 x 1 kernel with a stride of 2, and a 3 x 3 kernel with a stride of 2 and 1 pixels.

MobileNetV2. Submitted by (Sandler et al., 2018), and designed for mobile resource-constrained systems, which is an improvement of MobileNetV1. The network uses depth-wise separable convolutions which permit similar effects as convolutional layers, but decrease the single layer computations. The architecture of MobileNetV2 as shown in Figure A.6, Appendix A, structured from convolution layer with 32 filters followed with 17 bottleneck residual blocks to retain important information in the network, a regular 1 x 1 convolution, followed by a global average pooling layer and a ReLU activation as classification layer.

Table 3.1 describes some main features of the CNN networks, where the model size refers to the file size on actual disk after training the model on ImageNet dataset. Parameters refer to the total number of weights between the neuron connections. Depth refers to the topological depth of the network, includes activation layers, batch normalization layers etc.

Table 3.1. A summary of CNN models characteristics.

Model Name	Model Size	Total Layers	Parameters	Depth
Xception	88 MB	58	22,910,480	126
VGG19	528 MB	19	138,357,544	26
ResNet50	98 MB	50	25,636,712	-
MobileNetV2	14 MB	20	3,538,984	88
InceptionV3	92 MB	42	23,851,784	159
ResNet152V2	232 MB	152	60,380,648	-

3.4.2 CNNs Comparison Procedure

The input to individual CNN is an image (Y^s) from a CVC-clinic. The dataset $D = \{Y^s , L^s ; s = 1, ... S\}$ of the images which consists of S images that corresponds to the *labels* $L^s \in \{1, . . . , B\}$ for classification into B classes (binary). Each input dimension $(d \times d \times m)$ image, where $(d \times d)$ are spatial dimensions, and the (m) refer to the channels number in the image. The DeepCPC utilized RGB channel $(m=3)$, and these input images will be fed to the selected networks: ResNet50, Xception, MobileNetV2, ResNet152V2, InceptionV3 and VGG19.

3.4.3 Top Layers Configuration

The top layers from each CNN are removed to overcome the limitation of input image size. Each layer produces K output feature maps (bottleneck features) fed into the average pooling layer, and then to the proposed top layer. The proposed top layer, as shown in Figure 3.4, consists of two new Dense layers: (Dense-512 and Dense-2), and one dropout layer between the two FC layers. The dropout disables the detector nodes that have week features during the training. The activation functions used after the (FC-512) is rectifier function $f(x) = max (0, x)$, where x is the neuron input. The Dropout output will be the input for the next (FC-2). The softmax is being used as the output layer by the current top layer to predict a separate probability for each category: polyp or no-polyp.

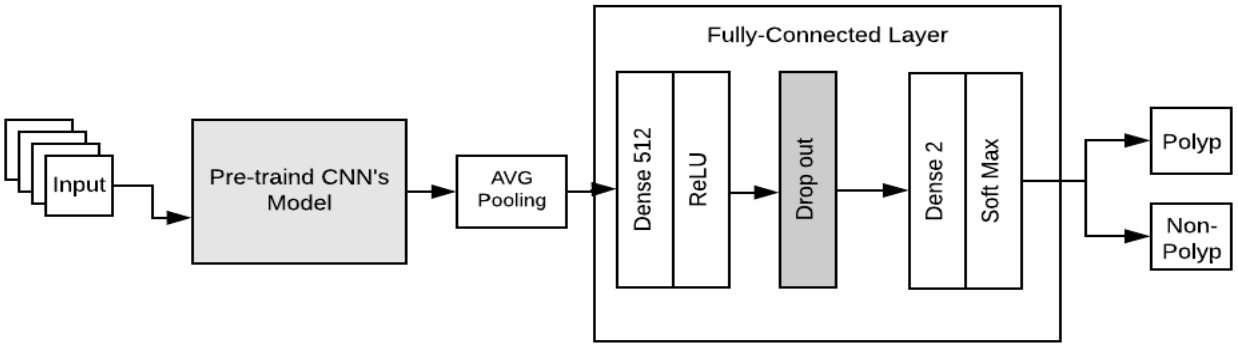


Figure 3.4. Transfer learning applied on CNN architecture.

3.4.4 Transfer learning and Fine Tuning

Since the numbers of polyp lesions in most of the images used are limited, therefore, transfer learning will be applied. At this stage, each individual CNN will be initialized with its weights that already trained on ImageNet, which has learned features that could be valuable for the classification. Moreover, freeze technique being used to the lower-level portion convolutional layers of each individual CNN, due to their ability to holding more generic features of the polyp. The training will be applying for the proposed fully connected layers with initialized weights loaded from a saved corresponding layer and fine-tuning the proposed classification system with the available CVC-Clinic dataset. Each single CNN will be referred to as following; FT-Xception, FT-VGG19, FT-ResNet50, FT-MobileNetV2, FT-InceptionV3, FT-ResNet152V2, where (FT) refer to Fine-Tuning.

3.5 Features Concatenation

After applying the transfer learning technique on the mentioned convolutional neural networks and feeding with polyp images as the input, a model evaluation is being applied on image features extracted from each individual convolutional neural network. In addition, according to the classification accuracy results from each single network, features concatenation is being performed by utilize two CNNs achieved the highest performance. As a result, a single image descriptor is formulated and used to represent each colorectal polyps and non-polyp image. The following subsection, present the procedure of features extraction, selection, and concatenation.

3.5.1 Features Extraction and Selection

Given two CNNs from the previous stage, and let the extracted features from the last convolutional layers in intermediate level denoted by the matrix ($X \in R^{a \times C}$) and ($Y \in R^{a \times D}$), where C and D represent the total number of feature maps that can differ according to the specific CNN architecture. The size of feature maps after the pooling layer is (a). For each matrix, the x_i and y_i is the i^{th} column that corresponds to one feature map. The output features are extracted from the activation functions of the last convolutional layer in the specific CNN model is:

$$R = \left(\sum_{i=1}^{c} \omega x_i + \beta\right) \odot \left(\sum_{i=1}^{d} \omega y_i + \beta\right) \tag{3.1}$$

Where $\odot$ is an element-wise product, ω are the weights for each feature map, and β represents the bias. The feature map (R) represents the bottleneck features.

3.5.2 Feature Concatenation Procedure

The feature maps, obtained from each CNN intermediate layers, are directly concatenated into one single vector that represents one image, as shown in Figure 3.5.

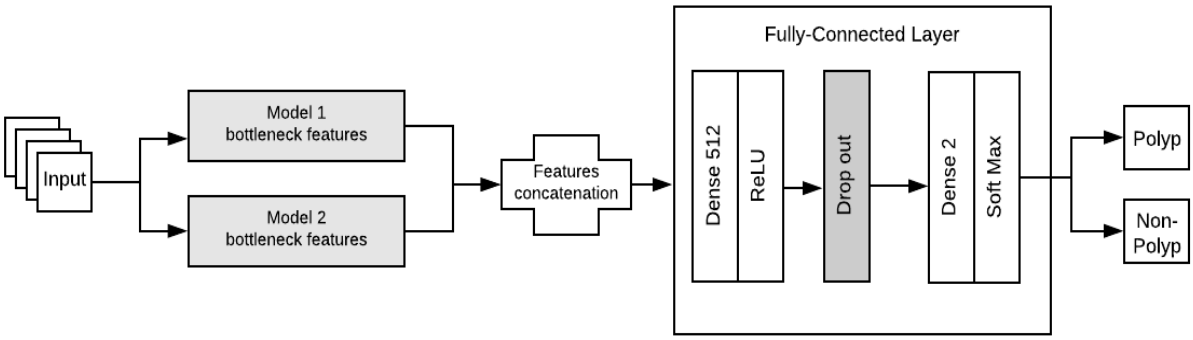


Figure 3.5. The proposed procedure of features concatenation.

Given two CNN models: the first model (K_1), and the second model (K_2). The combined features will be:

$$F = [K_1, K_2] \tag{3.2}$$

Where $K_1 \in R^{d \times d \times M1}$ refers to the bottleneck features map of model K_1, (d × d) denotes the width and height, and M1 is the number of channels. $K_2 \in R^{d \times d \times M2}$ is the bottleneck features map of model two. Then the combined features will be as:

$$F \in R^{d \times d \times (M1+M2)} \tag{3.3}$$

The concatenated features vector will be fed to the proposed fully connected layers, as discussed in section 3.4.3. Therefore, the complete model is fine-tuned on the polyps dataset via a number of end-to-end training epochs. Finally, a fully-trained deep model is obtained that will be used for polyps classification.

3.6 Model Performance Evaluation

This section, defines the evaluation metrics used to measure the performance DeepCPC model.

True Positive (TP): The number of positive samples that are identified correctly by the classifier means that sample falls in polyp class and classified as such.

False Positive (FP): The number of negative sample that are wrongly identified in a positive category, means that sample falls in non-polyp class, but classified as polyp class.

True Negative (TN): The number of negative sample that are identified correctly in its category. Samples are non-polyp class and classified as such.

False Negative (FN): The number of positive samples that are wrongly identified in another category means that the sample falls in polyp class, but classified as non-polyp class.

Confusion Matrix is a table utilized to describe the overall performance of the classification model on test data whose actual values are known. The relation between true positive, false positive, true negative and false negative are shown in Table 3.2.

Table 3.2. The calculation of TP, TN, FN, and FP.

		Actual Class	
		Polyp	Non-polyp
Predicated Class	Polyp	True positive	False positive
	Non-polyp	False negative	True negative

Recall (REC): Calculates the proportion of all true positive samples from cases that are actually positive. Also it referred to as sensitivity and true positive rate.

$$\text{REC} = \frac{True\ Positive}{Total\ Acutal\ Positive} = \frac{TP}{TP+FN} \tag{3.4}$$

Precision (PREC): Calculates the proportion of all true positive samples from cases that are predicated as positive.

$$\text{PREC} = \frac{True\ Positive}{Total\ Predicated\ Positive} = \frac{TP}{TP+FP} \tag{3.5}$$

Accuracy (ACC): Calculates the proportion of correctly classified samples.

$$\text{ACC} = \frac{True\ Positive+True\ Negative}{Number\ of\ Samples} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.6}$$

F1 score (F1): Another accuracy measure, also referred F-measure, utilized to seek the relation between precision and recall by counting the weighted average.

$$\text{F1 score} = 2 \times \frac{PREC \times REC}{PREC + REC} = \frac{2TP}{2TP+FP+FN} \tag{3.7}$$

ROC curve: The receiver operating characteristics is a two-dimensional graph in which created by plotting the false positive rate FPR on the x-axis against true positive rate TPR represents the y-axis at various threshold settings.

Specificity (SPEC): Calculates the proportion of all true negative samples from cases that are actually negative, also referred to false positive rate.

$$\text{SPEC} = \frac{True\ Negative}{Total\ Actual\ Negative} = \frac{TN}{TN+FP} \tag{3.8}$$

3.7 Summary

This chapter discussed the techniques used for the experiments and presented the DeepCPC architecture in details. It also presented the dataset in-depth utilized in the implementation. Then, it moved on to the CNN network employed in the DeepCPC including preprocessing, extracting features from CNN network, then showing the concatenation features technique applied on the best two CNN performance. Finally, it covered different performance matrices including confusion matrices, recall, precision, accuracy, F1-score, ROC-curve, and specificity. The next chapter will discuss the implementation details and experimental results.

Chapter 4

Implementation and Results

This chapter presents in-depth a discussion on the DeepCPC configurations, implementation, and experimental results for automatic polyp classification. Section 4.1 illustrates the environment configurations; Section 4.2 provides the software toolkits and hardware requirements; Section 4.3 presents the preparation procedure of the CVC-clinic dataset including patch extraction and data splitting; Section 4.4 illustrates the experimental scheme conducted on a set of deep CNN architectures to initialize DeepCPC, which highlights specific hyper-parameters and their performance on the transfer learning efficiency; Section 4.5 presents a detailed implementation of the DeepCPC including features extraction and formulation, and the results are discussed and evaluated using standard performance metrics with some comparisons; Section 4.6 summarizes this chapter.

4.1 Experiments Configuration

All the experiments were carried out on Co-laboratory model provided by Google. It is a cloud computing service that allows performing professional and advanced projects using Jupyter, which enables us to implement the entire project in Python. In addition to the provided virtual machine terminal, the hardware specifications allocated and used in this project are as follows:

- GPU Tesla P100-PCIE-16GB.
- 4 Intel(R) Xeon(R) CPU @ 2.20GHz.
- 15GB RAM.

4.2 Programming Language and Libraries

There are several deep learning resources available to use for developers, which made the implementation components and procedures efficient. Several toolkits and libraries are used after careful consideration and based on the unique requirements and time limitations of this work. The following subsections will introduce them.

4.2.1 Keras

Is a Python-based open Source application programming interface that uses either Theano or Tensorflow as backend (Team, 2020). It was designed to allow rapid innovation, allowing complete solutions to be more easily created and readable with the biggest range of CNN model.

4.2.2 TensorFlow

TensorFlow is a Python open source library created and released by Google (Tensorflow, 2020), for rapid numerical computing under an open source license from Apache 2.0. It is a base library that is used to develop Deep Learning models or to facilitate the process built on TensorFlow using any other wrapper library, such as Keras. It can operate on single CPU systems, mobile systems, GPUs, and distributed large-scale systems on centuries of machines.

4.2.3 Python for Data Science

Python is a language of high-level programming that has excellent library and community capabilities for data science applications. The libraries utilized in this work as the following:

- *MatPlotLib*: a popular multifunctional 2D and 3D plotting library.
- *NumPy*: a fundamental package for computational computing, defining arrays, matrices, and operations on these, which is used in Python's core scientific computing packages.
- *SciPy*: a domain toolboxes collection, and numerical algorithms.

4.3 Input Data Preparation

The classification performance of the DeepCPC model is influenced by the proper preparation of images dataset. CVC-Clinic DB dataset is used to evaluate DeepCPC, and the dataset includes images with sufficient diversity. Due to the limited amount and highly imbalanced types of images in CVC-Clinic dataset, a patch extraction and data augmentation are performed to provide more data samples utilized in the training phase. The next sub-sections introduce these processes thoroughly.

4.1.1 Patch Extraction

The CVC-Clinic dataset is imbalanced, means the dataset only represents one class of polyp image. As the DeepCPC requires binary classification (polyp, non-polyp), patch extraction has been utilized to create a balanced dataset from the original dataset. The methodology for patch extraction as follows:

Polyp patches: This process extracts the patch covers the whole polyp from every frame by eliminating the black margin that appears in the ground truth, and it extracts the corresponding section image from the original images, i.e. the white area visible in the ground truth.

Non-polyp patches: Extract the region that does not contain any part of the polyp from each frame by eliminating the white area in the ground Truth which corresponding with the original. Figure 4.1 illustrates the process of extracting polyp and non-polyp patches from CVC-Clinic DB.

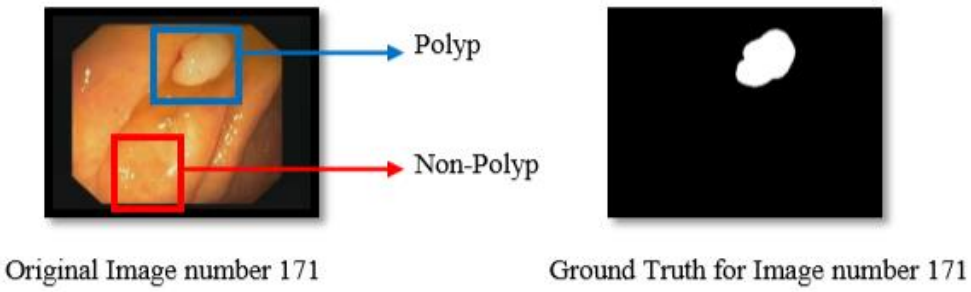


Figure 4.1. Illustrate patches extraction for Polyp and Non-Polyp.

The data augmentation techniques are being utilized with random rotations, horizontal and vertical flip, zoom-in, and zoom-out, therefore, the number of positive and negative samples can be artificially increased, and finally generating new balanced dataset with total of 1222 samples. For polyp 611 samples and for non-polyp 611 samples.

4.1.2 Dataset Split

The dataset of 1222 images are split into 80/20 split ratio, i.e. 80% for training and 20% for testing. Then, the training images are also split into training and validation datasets using 80/20 ratio, i.e. 80% for training and 20% for holdout validation.

Two different main classes used as targets in the training and testing phases, and they organized into folders that contain two subfolders titled 0 and 1, where 0 represents a group of non-polyp images and 1 is responsible for representing lesions classified as a polyp. Table 4.1 summarizes the splitting procedure of training, validation, and testing samples.

Table 4.1. Training-testing split details.

Method	Ratio (%)	Training	Validation	Testing
Split sample	80/20	781	196	245

4.4 Experiments of Model Initialization

As described in Section 3.4, the experiments conducting on a set of predefined CNN architecture to set up the first block of our architecture for polyps. The aim of experiments carried out here on each individual deep architecture is to figure out the best performing ones on the polyps dataset. Consequently, the performance of six distinct state-of-the-art CNNs are being examined under the same experimental configurations. These configurations are important to be carefully considered as they largely affect the overall accuracy of any trained model. The following subsection, presents the experimental setups made and used in each step.

4.4.1 Hyper-parameters Optimization

Each input image is resized to 150 × 150 × 3. All CNN networks are trained with an initial learning rate (η) 0.00001 using Adam optimizer (Opt). A batch size (Bs) of 64 is adopted in all the experiments of model initialization and transfer learning. Each individual network is trained for 100 epochs (Te). Table 4.2 summarizes these initialized hyperparameters.

Table 4.2.Hyper parameters used in model initialization.

Training parameter	Value
T e	100
Bs	64
Is	150 × 150 × 3
η	0.00001
Opt	Adam: η =0.001

4.4.2 Transfer Learning

Firstly, a new top layer is implemented to replace the existing fully connected layers in order to fine-tune them on the new domain, i.e. polyp medical images. Each individual pre-trained network is used as feature extractor and only the new top layers are trained again on CVC-Clinic

dataset. Furthermore, transfer the learned ImageNet weights as initial weights, and fine-tune the customized model with the new top layer through a complete end-to-end training, i.e. forward and back-propagation procedures. The classification performance of all individual optimized CNN models is evaluated using 245 testing images (20% of dataset).

4.4.3 Experimental Results of Model Initialization

Table 4.3 shows the detailed results for each pre-trained CNN performance that has been modified and trained on the CVC-Clinic dataset. Figure 4.3 shows the Precision-Recall curve for each model. As shown, the highest overall accuracy in the training and validation phases is achieved by the FT-Xception model with accuracy of (97.5%) followed by the FT-MobileNetV2 with accuracy of (97.1%). The rest of the networks are still performing well but less than FT-Xception and FT-MobileNetV2.

Moreover, the modified FT-Xception and FT-MobileNetv2 are also performing efficiently in terms of F1 score measures by reporting (97%) followed by FT-InceptionV3 that achieved (96%). The FT-ResNet50 and FT-ResNet152V2 achieved the lowest F1 scores compared to other architectures with results (92%) and (70%), respectively. It is worth to mention that FT-VGG-19 shows the lowest accuracy results less than (21%), but it gains a recall of 45.0% due to the fact this network predicted one class only.

Table 4.3. Evaluation of CNN's models using CVC-clinic DB images as input.

Methods	**FT-Xception**	**FT-MobileNetV2**	**FT-InceptionV3**	**FT-ResNet50**	**FT-ResNet152V2**	**FT-VGG-19**
Precision	0.97	0.97	0.96	0.92	0.81	0.21
Recall	0.97	0.97	0.96	0.90	0.73	0.45
F1-Score	0.97	0.97	0.96	0.92	0.70	0.28
Accuracy	0.976	0.971	0.956	0.927	0.727	0.453
Test Results						
TP	105	109	106	103	45	111
TN	134	129	133	124	133	0
FP	6	2	5	8	66	0
FN	0	5	1	10	1	134

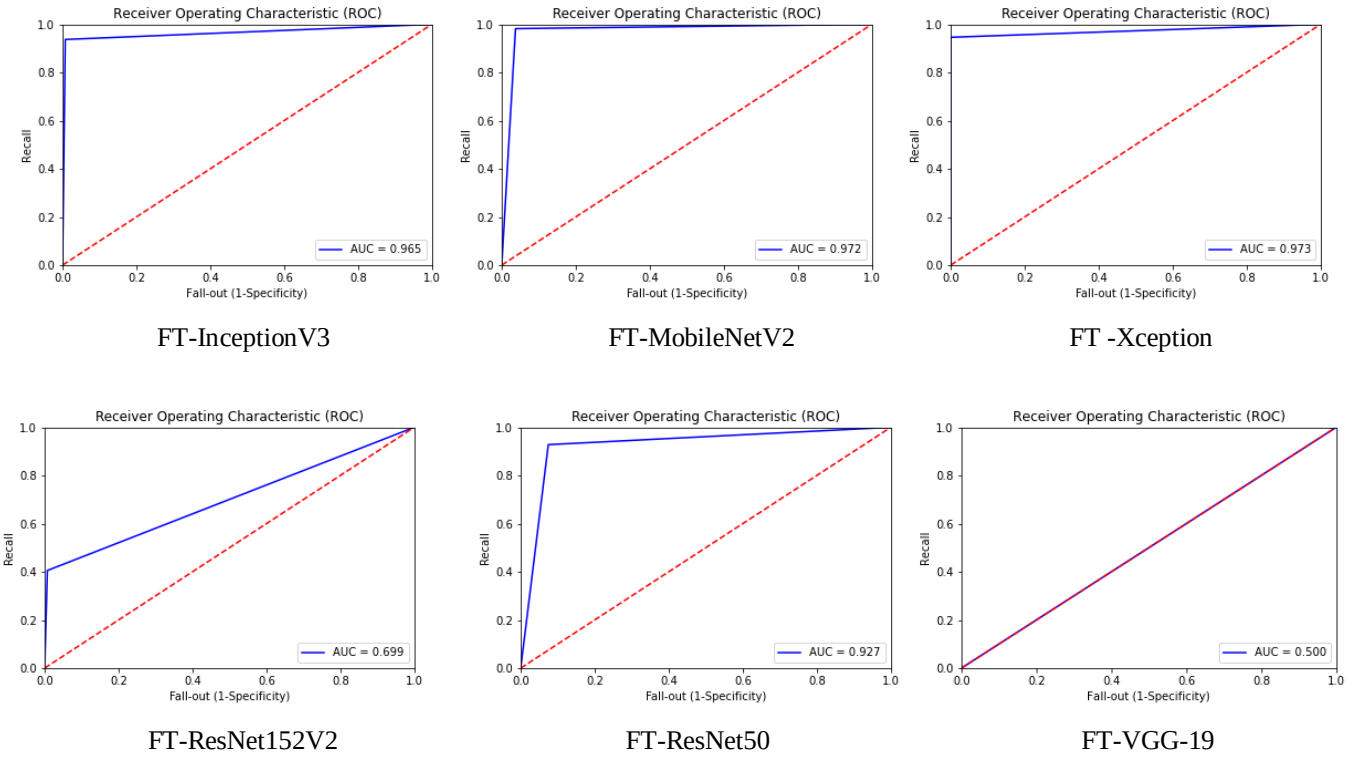


Figure 4.2.ROC Curves for Polyp Classification.

Obviously, the transfer learning scheme provided by the proposed architecture shows a noticeable impact on the accuracy results of the pre-trained CNN models. This initial finding is important to highlight how different convolutional features of polyps images do act and perform on this dataset when the transfer the learning is performed from general purpose domain, i.e. ImageNet, to a new different domain, i.e. medical images.

Moreover, the size of extracted features in each CNN model is reduced to 512-dimension vector that represents a generic image descriptor, which is a compact size compared to the original one. The full-size features are projected into a low-dimensional data space using the dense layer added to the top fully connected layers of the modified architecture. This reduction procedure confirms that the modified CNN architecture is able to achieve high accuracy even with compact image descriptors. Additionally, this will reduce the number of parameters to be learned, the training time, and the required memory to store the model and its parameters into the actual disk.

The performance impact of such architecture is beneficial in real-time applications and large-scale image repositories.

The next optimization approach in this thesis is enhancing the classification accuracy further using the concatenation procedure, which relies on utilizing the highest performance from two CNN architectures: (FT-Xception and FT-MobileNetV2).

4.5 Features Concatenation Experiments

This section, presents in details the stage three of DeepCPC model, as discussed earlier in Section 3.5. concatenating approach is utilized for the convolutional features extracted from the fine-tuned deep CNNs that achieved the highest performance results, which in turn provides a new single model that generates one fused image descriptor for any given polyps image. The following subsections introduce and discuss the configurations set for the input architectures (FT-Xception and FT-MobileNetV2), hyperparameters initialization, and experimental results of the optimized deep model.

4.5.1 FT-Xception Setups

The FT-Xception architecture, as mentioned in Section3.4.1, consists of 36 convolutional layers that are structured in three flows. The exit flow consists of 4 separable conventional layers with kernel size 3x3, and one conventional layer with kernel size 1x1 and stride 2x2, and followed by an average pooling layer, Figure A.6, Appendix A. From the last convolutional layer of the FT-Xception the features are being extracted, and a 2048-dimensional features vector is used in feature concatenation without considering the fully connected layer (bottleneck features).

4.5.2 FT-MobileNetV2 Setups

The FT-MobileNetV2 architecture achieved the second higher performance in the comparison stage and was selected for performing features concatenation with the FT-Xception network. MobileNetV2 consists of sixteen blocks. The Final flow structured with a conventional layer, batch normalizes and ReLU Activation that repeated three times followed by a conventional layer 1 x 1. The features from the last convolutional layer of the FT-MobileNetV2 are being extracted, and a 1280-dimensional features vector is used in feature concatenation without considering the fully connected layer (bottleneck features).

4.5.3 Hyper-parameter Initialization

This model was trained end-to-end for 100 epochs using a learning rate of 0.00001; a batch size of 64, an image size is of 150×150×3, and Adam optimizer with learning rate 0.001.

4.5.4 Final Model Fine-tuning

The features extracted by the two pre-trained CNN models are utilized, i.e. 2048-features vector from FT-Xception and 1280-features vector from FT-MobileNetV2. Concatenation technique is performed to the extracted features to form a 3328-dimensional features vector. The resulting vector's size is undesirable especially in online training systems and large-scale repositories; therefore, this vector is fed into a dense layer to reduce the size into only 512-features vector.

Since the concatenated convolutional features represent various details and characteristics of polyp images, this process helps in exploring the underlying image details and producing more inclusive depiction of discriminative features compared to the descriptions provided by any individual CNN model.

Finally, the customized fully-connected layers is being added, including the softmax layer classification, so the final image descriptor will be fed to this top part of DeepCPC and the performance will be reported and evaluated using a set of standard metrics.

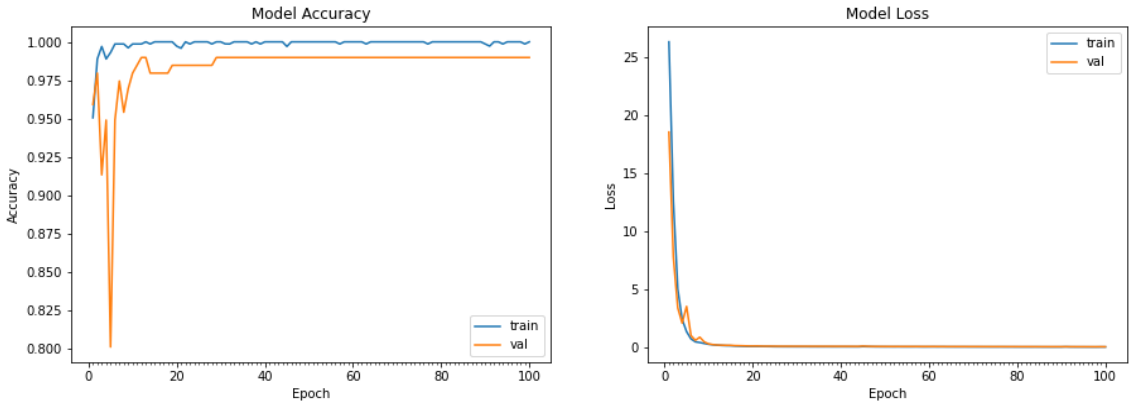
4.5.5 Results and Discussion

All experiments are conducted under the same defined setups on polyp vs non-polyp training images using the final model and its concatenated generic descriptor. The model performance is evaluated in terms of precision, recall, F1-score, accuracy, confusion matrix, training speed, and memory size. The performance results are shown in Figure 4.3. The results confirm the high capability of the DeepCPC in predicting non-polyp images, which reports 0.97, 1.00, and 0.99 scores of precision, recall, and F1-score, respectively. Similarly, achieves high accuracy in predicting polyp images by reporting scores of 1.00, 0.96, and 0.98 precision, recall, and F1-score, respectively.

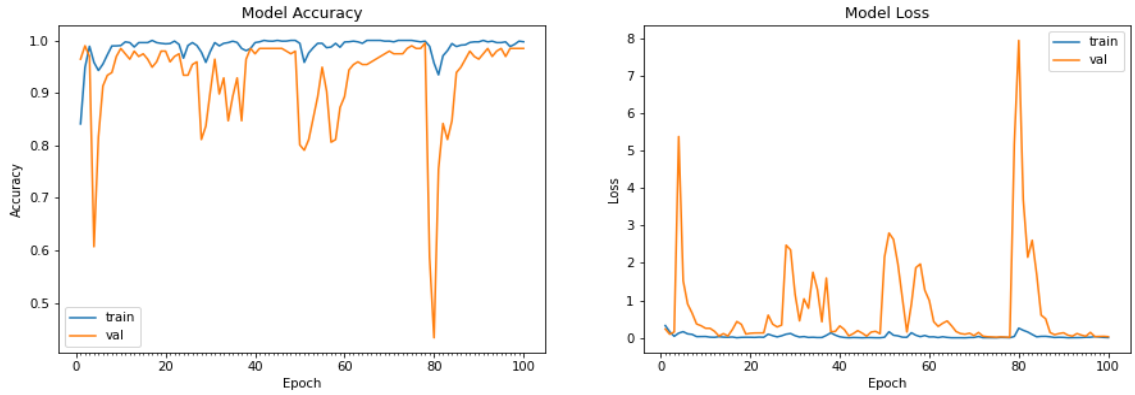
```
              precision    recall  f1-score   support

  Non-polyps       0.97      1.00      0.99       134
      polyps       1.00      0.96      0.98       111

    accuracy                           0.98       245
   macro avg       0.99      0.98      0.98       245
weighted avg       0.98      0.98      0.98       245

245/245 [==============================] - 1s 3ms/step
196/196 [==============================] - 0s 3ms/step
Validation: accuracy = 0.994898  ;  loss_v = 0.086271
Test: accuracy = 0.983673  ;  loss = 0.183429
```

Figure 4.3. The performance results using concatenation features.

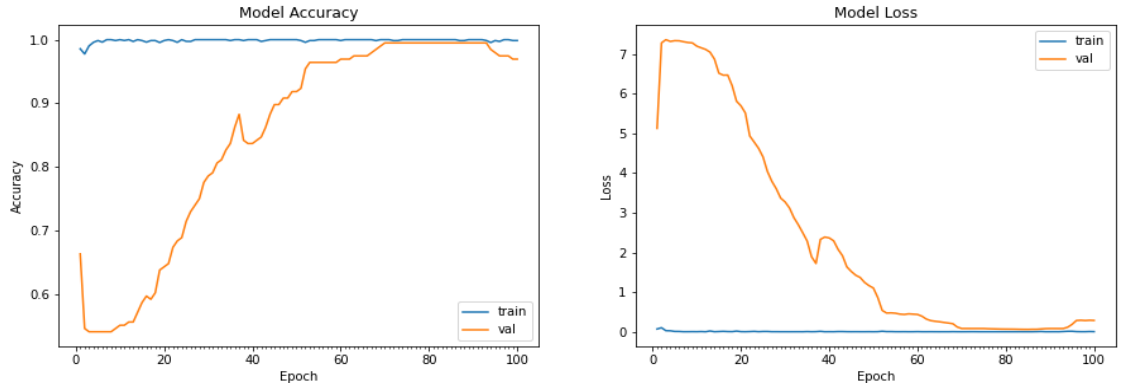
Figure 4.4(a) compares the learning curve of the DeepCPC using two concatenated networks. It can be easily observed that the loss rate of the DeepCPC is smoothly reduced to a lower value after epoch 5 while the accuracy increases smoothly after epoch 30. For FT-Xceptionin Figure 4.4(b), and FT-MobileNetV2 in Figure 4.4(c), there is a fluctuation in the training loss and accuracy throughout the training and validation procedures. This also confirms that the DeepCPC overcomes any possible overfitting or underfitting issues.



(a) Accuracy of network based on concatenated features.



(a) Accuracy of FT-Xception.



(c) Accuracy of FT-MobileNetV2.

Figure 4.4. Training and validation accuracy results.

Furthermore, the confusion matrix showed in Figure 4.5illustrates the number of misclassified images per classes. It can be observed that the DeepCPC classifies both types of images correctly. On one hand, the DeepCPC model makes a few misclassifications between polyp and non-polyp classes, i.e. only 2 misclassified images out of 111 for polyp class and 2 misclassified images out of 134 for non-polyp class. On the other hand, the confusion matrix of FT-Xception in Figure 4.6(a) shows 6 misclassified images out of 111 of polyp class but it is performing very well in predicting the non-polyp class. Also, the confusion matrix of FT-MobileNetV2 in Figure 4.6(b) shows that it performs similarly as the DeepCPC for polyp class with 2 misclassified images out of 111 but with 5 misclassified images out of 134 for non-polyp class.

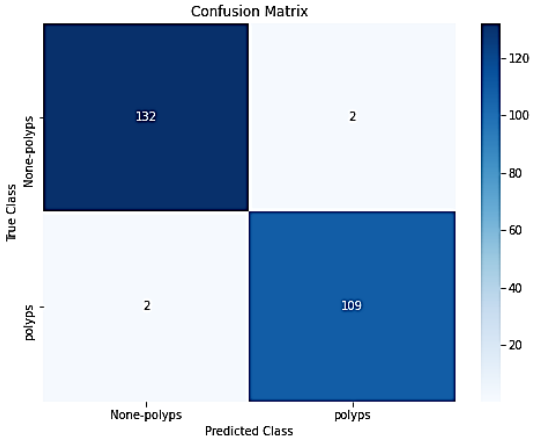
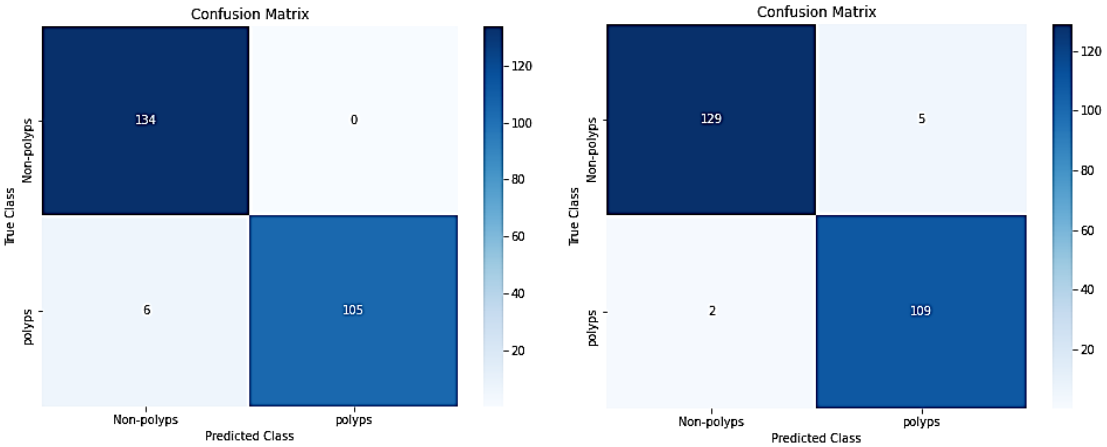


Figure 4.5.Confusion matrix of DeepCPC.



(a) FT-Xception (b) FT-MobileNetV2

Figure 4.6. Confusion matrix of FT-Xceptionm and FT-Mobile NetV2.

Finally, Table 4.4 summarizes and compares the performance of the classification results of single fine-tuned CNN models with DeepCPC. It shows the obtained average accuracy, precision, recall, and f1-score of the classifications. It is noted that the DeepCPC model performs better than single fine-tuned CNN by which the accuracy improved to 98.4% over the FT-Xception of 97.6% and FT-MobileNetV2 of 97.1%.

Table 4.4. Performance comparisons of the DeepCPC.

Methods	**FT-Xception**	**FT-MobileNetV2**	**FT-Concatenated**
Precision	0.97	0.97	0.98
Recall	0.97	0.97	0.98
F1-Score	0.97	0.97	0.98
Accuracy	0.976	0.971	0.984

4.6 Summary

- This chapter, various experiments has been conducting with a thorough analysis of results, described the environment and libraries used for the experiments and the evaluation metrics to measure our model performance. The model initialized with six pre-trained CNN, their performance is examined, and how transfer learning with fine-tuning affected the classification rates is discussed. The result in the implementation shows that FT-Xception and FT-MobileNetV2 outperform the other CNN. Finally, provide a features concatenation scheme based on the extracted bottleneck features from FT-Xception and FT-MobileNetV2 to enhance the model performance. The DeepCPC showed improvements in the overall performance, and it was able to improve the sensitivity from 97% to above 98%. The next chapter will present the research conclusion and future work.

Chapter 5

Conclusion and Future Work

5.1 Conclusions

In this thesis, various techniques and solutions has been investigated for the automatic classification of polyps in colonoscopy images. The aim of this thesis is to develop a deep learning model for classifying colorectal polyps based on discriminative features extracted from deep convolutional neural networks in order to achieve automated disease diagnosis. Firstly, a background of the colorectal polyp and its surveillance tests has been studied, in additions, discussed work on topics related to automatic polyp classification in colon images. For machine learning, an overview of different learning classes is introduced such as supervised, unsupervised, and reinforcement learning, then a discussion of convolutional neural network architecture and its parameters are illustrated. As for benchmark CNN architectures, the most popular CNN networks for image classification tasks are studied, such as VGGNet, GoogleNet, ResNet, MobileNet, and Xception, which can be utilized in the processing pipelines of health care fields.

Most importantly, a deep learning model is proposed (DeepCPC), which consists of three stages. The first stages present the image preprocessing including techniques used such as patch extraction and data augmentation to classify the original dataset into two types, i.e. polyp and non-polyp, because the original CVC-Clinic dataset contains polyp image only. The second stage presents the model initialization by modifying and comparing the pre-trained CNN networks, adapting transfer learning scheme, and fine tuning the models with the proposed fully connected layer. The final stage presents the core design of DeepCPC by adapting features concatenation from the two pre-trained networks modified that achieved the highest performance in the previous stage. According to the comparisons results of six CNN models, the FT-Xception and FT-MobileNetV2 have shown highest performance with average accuracy of 97.5% and 97.1%, respectively.

Finally, an overall 98.4% classification accuracy, 98% of precision, 98% of sensitivity, and 98% of f1-score are achieved by using features concatenation technique with the proposed fully connected layer and optimized hyper-parameters, then fine-tuning DeepCPC on polyps images through a complete end-to-end training procedure, which outperformed the single transfer learning classification methods in each defined performance metric. The DeepCPC structure is also flexible and dynamic, such that it can be effectively expanded in the future to incorporate the classification in other forms of diseases.

5.2 Future Work

A number of improvements could be applied while implementing a deep learning-based solution of this kind in the health care sector. The presented results are promising but several improvements could be considered in the future as follows:

- Increasing the data set could yield better performance for classifying colorectal polyp. A larger data set also introduces the possibility of including more classes, enabling the classification of other polyp types as well. Including serrated sessile, pedunculated, and tubular, etc. would thereby widen the area of use. Therefore, it is a necessity to further collect a greater number of colonoscopy images in order to re-evaluate, qualify and respectively confirm the results of this work, further performing the optimization of the DeepCPC model to be able to classify all different polyp types in future.
- Investigate the influence of different approaches for features dimensionality reduction such as principal component analysis (Pearson, 1901), or random projection (Johnson & Lindenstrauss, 1984).
- Other types of local features can be utilized with convolutional features including fisher vectors (Jaakkola & Haussler, 1998) and scale invariant feature transform (Lowe, 2004).

Bibliography

The American Cancer Society medical and editorial content team. (2018, 02 21). *What Is Colorectal Cancer?* (cancer.org) Retrieved 10 01, 2019, from https://www.cancer.org/cancer/colon-rectal-cancer/about/what-is-colorectal-cancer.html

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2016). Global patterns and trends in colorectal cancer incidence and mortality. *Gut, 66*. doi:10.1136/gutjnl-2015-310912

Awad, M., & Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers.* Berkeley, CA: Apress.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 2481-2495. doi:10.1109/TPAMI.2016.2644615

Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., & Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics, 43*, 99-111 .

Bernal, J., Sánchez, J., & Vilariño, F. (2012). Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition, 45*, 3166-3182. doi:0.1016/j.patcog.2012.03.002

Bernal, J., Tajkbaksh, N., Sánchez, F.J., Matuszewski, B., Chen H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., Pogorelov, K., Choi, S., Debard, Q., Maier-Hein, L., Speidel, S., Stoyanov, D., Brandao, P., Cordova, H., Sánchez-Montes, C. (2017). Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Transactions on Medical Imaging*.

Burkov, A. (2019). *The hundred-page machine learning book.* Andriy Burkov.

Byrne, M. F., Chapados, N., Soudan, F., Oertel, C., Linares Pérez, M., Kelly, R., … Rex, D. K. (2017). Real-time differentiation of adenomatous and hyperplastic diminutive colorectal

polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, 94–100. doi:0.1136/gutjnl-2017-314547

Chen, J., Milot, L., Cheung, H. M. C., & Martel, A. L. (2019). Chen, J., Milot, L., Cheung, H. M. C., & Martel, A. L. (2019). Unsupervised Clustering of Quantitative Imaging Phenotypes Using Autoencoder and Gaussian Mixture Model. *Lecture Notes in Computer Science Medical Image Computing and Computer Assisted Intervention – MICCAI*, 575–582. doi:10.1007/978-3-030-32251-9_63

Dekker, E., & Rex, D. K. (2018). Advances in CRC prevention: screening and surveillance. *Gastroenterology, 154*. doi:10.1053/j.gastro.2018.01.069

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition*, 248–255.

Ferlay, J. , Colombet, M. , Soerjomataram, I. , Mathers, C. , Parkin, D. , Piñeros, M. , Znaor, A. and Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018:GLOBOCAN sources and methods. *International Journal of Cancer, 144*, 1941-1953. doi:10.1002/ijc.31937

Groff, R. J., Nash, R., & Ahnen, D. J. (2008). Significance of serrated polyps of the colon. *Current gastroenterology reports*, 490–498. doi:10.1007/s11894-008-0090-z

Guo, Y., & Matuszewski, B. (2019). GIANA Polyp Segmentation with Fully Convolutional Dilation Neural Networks. *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*.

He, K., Zhang, X., Ren, S., & Sun, J. . (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hilsden RJ, Heitman SJ, Mizrahi B, Narod SA, Goshen R. . (2018). Prediction of findings at screening colonoscopy using a machine learning algorithm based on complete blood counts (ColonFlag). *PLoS One*, 1-9. doi:10.1371

Horváth, A., Spindler, S., Szalay, M., & Rácz, I. (2016). Preprocessing Endoscopic Images of Colorectal Polyps. *Acta Technica Jaurinensis, 9*, 65. doi:0.14513/actatechjaur.v9.n1.397

Huang, Y., Gong, W., Su, B., Zhi, F., Liu, S., & Jiang, B. (2012). Risk and Cause of Interval Colorectal Cancer after Colonoscopic Polypectomy. *Digestion, 86*, 148-154. doi:10.1159/000338680

Irshad, H., Veillard, A., Roux, L., Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: a reviewcurrent status and future potential. *IEEE reviews in biomedical engineering, 7*, 97-114.

Jaakkola, T. & Haussler, D. (1998). Exploiting Generative Models in Discriminative Classifiers. In *In Advances in Neural Information Processing Systems 11* (pp. 487--493). MIT Press.

Jerebko, A. K., Malley, J. D., Franaszek, M., & Summers, R. M. (2003). Computer-aided polyp detection in CT colonography using an ensemble of support vector machines. *International Congress Series, 1256*, 1019-1024. doi:10.1016/S0531-5131(03)00314-5

Johnson, W. B., & Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. *Conference on Modern Analysis and Probability Contemporary Mathematics*, 189–206. doi:10.1090/conm/026/737400

kaggle. (2013). *Challenges in Representation Learning: Facial Expression Recognition Challenge*. Retrieved 04 01, 2019, from https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

Kahi, C. J. (2015). How does the serrated polyp pathway alter CRC screening and surveillance? *Digestive diseases and sciences, 60*. doi:10.1007/s10620-014-3449-z

Komeda Y, Handa H, Matsui R, Kashida H, Watanabe T, Sakurai T, et al. (2017). Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncology*. doi:10.1159/000481227

Komeda, Y., Handa, H., Matsui, R., Kashida, H., Watanabe, T., Sakurai, T., & Kudo, M. (2019). Computer-Aided Diagnosis (Cad) Based On Convolutional Neural Network (Cnn) System Using Artificial Intelligence (Ai) For Colorectal Polyp Classification. *ESGE Days 2019*.

Kopelman Y, Gal O, Jacob H, Siersema P, Cohen A, Eliakim R, Zaltshendler M, Zur D. (2019). Automated Polyp Detection System in Colonoscopy Using Deep Learning and Image Processing Techniques. *Journal of Gastroenterology and Its Complications, 3*(1).

Li, P., Chan, K. L., Krishnan, S., & Gao, Y. (2004). Detecting abnormal regions in colonoscopic images by patch-based classifier ensemble. *Proceedings of the 17th International Conference on Pattern Recognition, 3*.

Li, Q., Yang, G., Chen, Z., Huang, B., Chen, L., Xu, D., … Wang, T. (2017). Colorectal polyp segmentation using a fully convolutional neural network. *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI),* 1-5. doi:10.1109/CISP-BMEI.2017.8301980

Lieberman, D. A., Rex, D. K., Winawer, S. J., Giardiello, F. M., Johnson, D. A., & Levin, T. R. (2012). Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology, 143,* 844-857. doi:10.1053/j.gastro.2012.06.001

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 91-110. doi:10.1023/b:visi.0000029664.99615.94

Manivannan, S., Wang, R., Trucco, E., & Hood, A. (2013). Automatic normal-abnormal video frame classification for colonoscopy. *IEEE 10th International Symposium on Biomedical Imaging*, 644-647.

Michaeli, T., Eldar, Y. C., & Sapiro, G. . (2012). Semi-supervised single- and multi-domain regression with multi-domain training. *Information and Inference*. doi: 10.1093/imaiai/ias003

Mo, X., Tao, K., Wang, Q., & Wang, G. (2018). An Efficient Approach for Polyps Detection in Endoscopic Videos Based on Faster R-CNN. *2018 24th International Conference on Pattern Recognition (ICPR), 2018*(1809.01263v1).

Mohammed, A., Farup, I., Pedersen, M., Hovde, Ø., & Yayilgan, S. Y. (2018). Stochastic Capsule Endoscopy Image Enhancement. *Journal of Imaging, 4*(10.3390/jimaging4060075).

Mori Y, Kudo SE, Berzin TM, Misawa M, Takeda K. (2017). Computer-aided diagnosis for colonoscopy. *Endoscopy, 49(8),* 813–819. doi:10.1055/s-0043-109430

Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., & Sivaprakasam, M. (2019). Joint shape learning and segmentation for medical images using a minimalistic deep network. *arXiv preprint, 1*(1901.08824), 14.

Nguyen, N.-Q., & Lee, S.-W. (2019). Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network. *IEEE Access, 7*, 33795-33808.

Nicholls, J. G., & Kuffler, S. W. . (2018). *From neuron to brain.* Sunderland: MA: Sinauer Associates, Inc.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 559–572. doi:10.1080/14786440109462720

Pedro N. Figueired, Isabel N. Figueiredo, Luís Pinto, Sunil Kumar, Yen-Hsi Richard Tsai, Alexander V. Mamonov. (2019). Polyp detection with computer-aided diagnosis in white light colonoscopy: comparison of three different methods. (E209–E215).

Person. (2015, Dec 9). *10 Breakthrough Technologies*. (MIT Technology Review) Retrieved 12 1, 2019, from https://www.technologyreview.com/lists/technologies/2013/

Pogorelov, K., Suman, S., Azmadi Hussin, F., Saeed Malik, A., Ostroukhova, O., Riegler, M., … Goh, K. L. (2019). Bleeding detection in wireless capsule endoscopy videos - Color versus texture features. *Journal of applied clinical medical physics, 20(8)*, 141–154. doi:10.1002/acm2.12662

Qadir, H. A., Balasingham, I., Solhusvik, J., Bergsland, J., Aabakken, L., & Shin, Y. (2019). Improving Automatic Polyp Detection Using CNN by Exploiting Temporal Dependency in Colonoscopy Video. *IEEE Journal of Biomedical and Health Informatics*. doi:10.1109/JBHI.2019.2907434

Qaiser, T., Tsang, Y.-W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D., & Rajpoot, N. (2018). Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis, 1*(1805.03699).

Rex, D., Cutler, C., Lemmel, G., Rahmani, E., Clark, D., Helper, D., … Mark, D. (1997). Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology, 112*, 24-28.

Ribeiro, E., Uhl, A., Häfner, M. (2016). Colonic Polyp Classification with Convolutional Neural Networks. *In Proceedings of the IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, 253–258.

Ribeiro, Eduardo, et al. . (2016). Colonic Polyp Classification with Convolutional Neural Networks. *IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. doi:10.1109/cbms.2016.39

Rouse, M. (2014). *RSA algorithm (Rivest-Shamir-Adleman)*. Retrieved 11 16, 2017, from http://searchsecurity.techtarget.com/definition/RSA

S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and. (2003). Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE transactions on information technology in biomedicine*, 141-152. doi:10/1109

Sarkar, D., Bali, R., & Ghosh, T. (2018). *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras.* Birmingham: Packt Publishing.

Shalev-Shwartz, S., & Ben-David, S. (2017). *Understanding machine learning: from theory to algorithms.* Cambridge : Cambridge University Press.

Silva.J, Histace.A, Romain.O, Dray.X, Granado.B . (2014). Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *Int. J. Comput. Assist. Radiol. Surg, 9*, 283–293.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv 1409.1556*.

Srivastava, Ashok N. and Han, Jiawei. (2011). *Machine Learning and Knowledge Discovery for Engineering Systems Health Management* (1st ed.). Chapman & Hall/CRC.

Stehle, T., Auer, R., Gross, S., Behrens, A., Wulff, J., Aach, T., … Tischendorf, J. (2009). Classification of colon polyps in NBI endoscopy using vascularization features. *Medical Imaging 2009: Computer-Aided Diagnosis*. doi: doi: 10.1117/12.808103

Summers, R. M., Yao, J., Pickhardt, P. J., Franaszek, M., Bitter, I., Brickman, D., … Choi, J. R. (2005). Computed Tomographic Virtual Colonoscopy Computer-Aided Polyp Detection in a Screening Population. *Gastroenterology, 129*(6). doi:10.1053/j.gastro.2005.08.054

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2015.7298594

T. Tamaki et al.,. (2013). Computer-aided colorectal tumor classification in NBI endoscopy using local features. *Medical Image Analysis*. doi:10.1016/j.media.2012.08.003

Tajbakhsh, N., Gurudu, S. R., & Liang, J. (2015). Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. *IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015*. doi:10.1109/ISBI.2015.7163821

Takamatsu, M., Yamamoto, N., Kawachi, H., Chino, A., Saito, S., Ueno, M., … Takeuchi, K. (2019). Prediction of early colorectal cancer metastasis by machine learning using digital slide images. *Computer Methods and Programs in Biomedicine*. doi:10.1016/j.cmpb.2019.06.022

Team, K. (2020, 4 20). *Keras documentation: About Keras.* Retrieved from keras: https://keras.io/about/

Tensorflow, t. (2020, 04 5). *Why TensorFlow*. Retrieved from TensorFlow: https://www.tensorflow.org/about

Tian, Y., Pu, L. Z., Singh, R., Burt, A. D., & Carneiro, G. (2019). One-Stage Five-Class Polyp Detection and Classification. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019*(10.1109/ISBI.2019.8759521).

Urban, G.; Tripathi, P.; Alkayali, T.; Mittal, M.; Jalali, F.; Karnes, W.; Baldi, P. (2018). Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology, 155*(4), 1069–1078. doi:10.1053/j.gastro.2018.06.037

Vázquez, D., Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., López, A. M., Romero, A., … Courville, A. (2017). A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *Journal of Healthcare Engineering, 2017*(4037190).

Wang, L., Chen, R., Wang, S., Zeng, N., Huang, X., & Liu, C. (2019). Nested Dilation Network (NDN) for Multi-Task Medical Image Segmentation. *IEEE Access*(10.1109), 44676–44685.

Wang, P., Berzin, T. M., Glissen Brown, J. R., Bharadwaj, S., Becq, A., Xiao, X., … Liu, X. (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*.

Wimmer, G., Vecsei, A., & Uhl, A. (2016). CNN transfer learning for the automated diagnosis of celiac disease. *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. doi: doi: 10.1109/ipta.2016.7821020

Zhang, R.; Zheng, Y.; Mak, T.W.C.; Yu, R.; Wong, S.H.; Lau, J.Y.W.; Poon, C.C.Y. (2017). Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed. Health Inf, 21*, 41–47.

Zhang, X., Chen, F., Yu, T., An, J., Huang, Z., Liu, J., … Si, J. (2019). Real-time gastric polyp detection using convolutional neural networks. *Plos One, 2019*.

Zheng, M., Krishnan, S., & Tjoa, M. (2005). A fusion-based clinical decision support for disease diagnosis from endoscopic images. *Computers in Biology and Medicine, 35*(3), 259-274.

Zhou, S. K., Greenspan, H., & Shen, D. (2017). Deep Learning for Medical Image Analysis. London: Elsevier/Academic Press.

Zufferey, T., Ulbig, A., Koch, S., & Hug, G. . (2018). Unsupervised Learning Methods for Power System Data Analysis. *Big Data Application in Power Systems*, 107–124. doi:10.1016/b978-0-12-811968-6.00006-1

Appendix A: Specifications of Pretrained CNNs Architectures

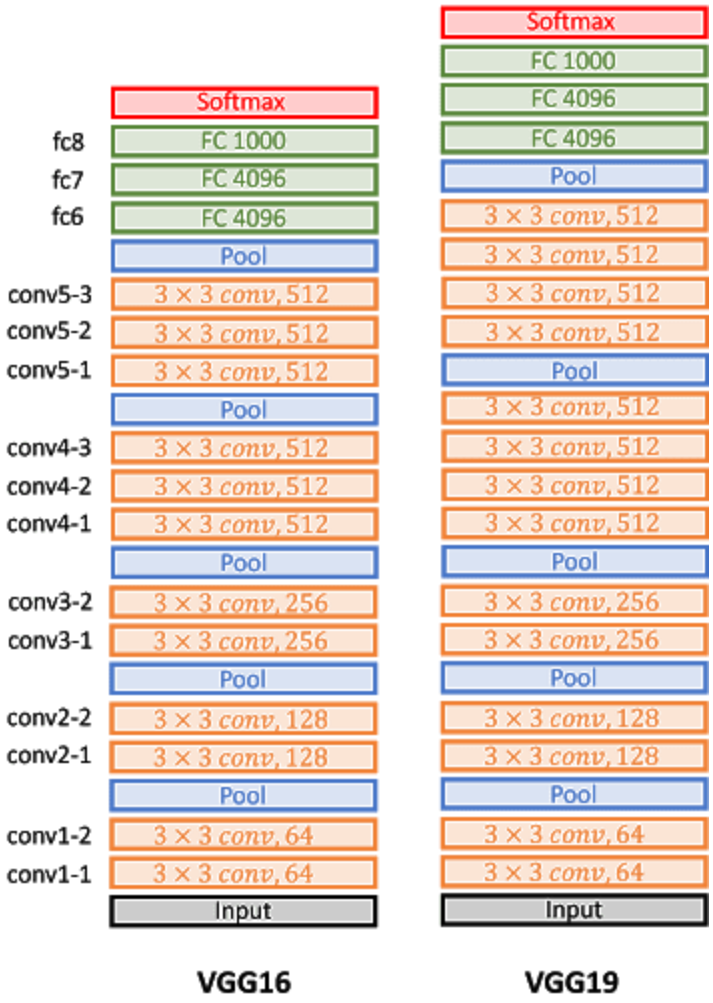


Figure A.1.VGG-16 and VGG-19 Architecture[10].

[10]Image fromdatahacher website available at: http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/

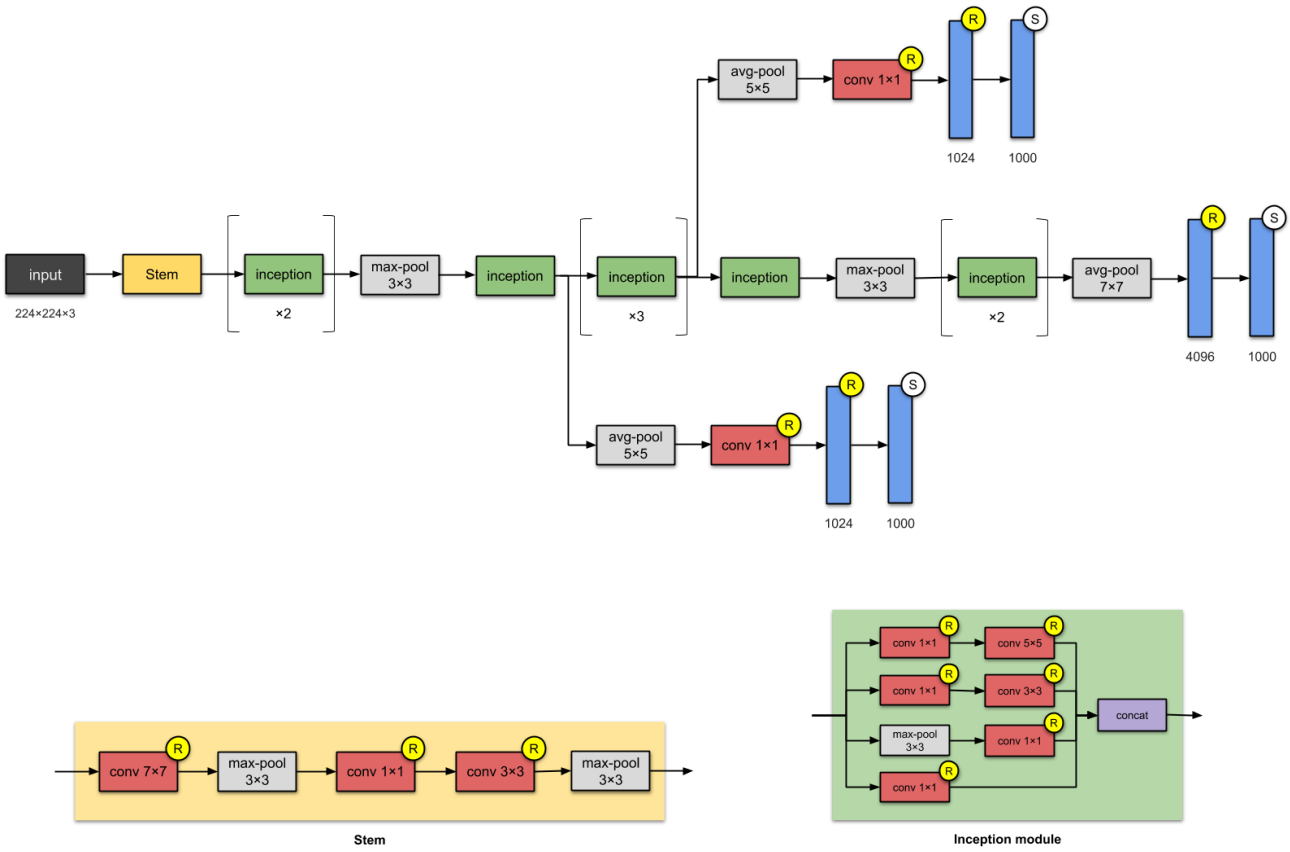


Figure A.2.Inception-V1 Architecture.

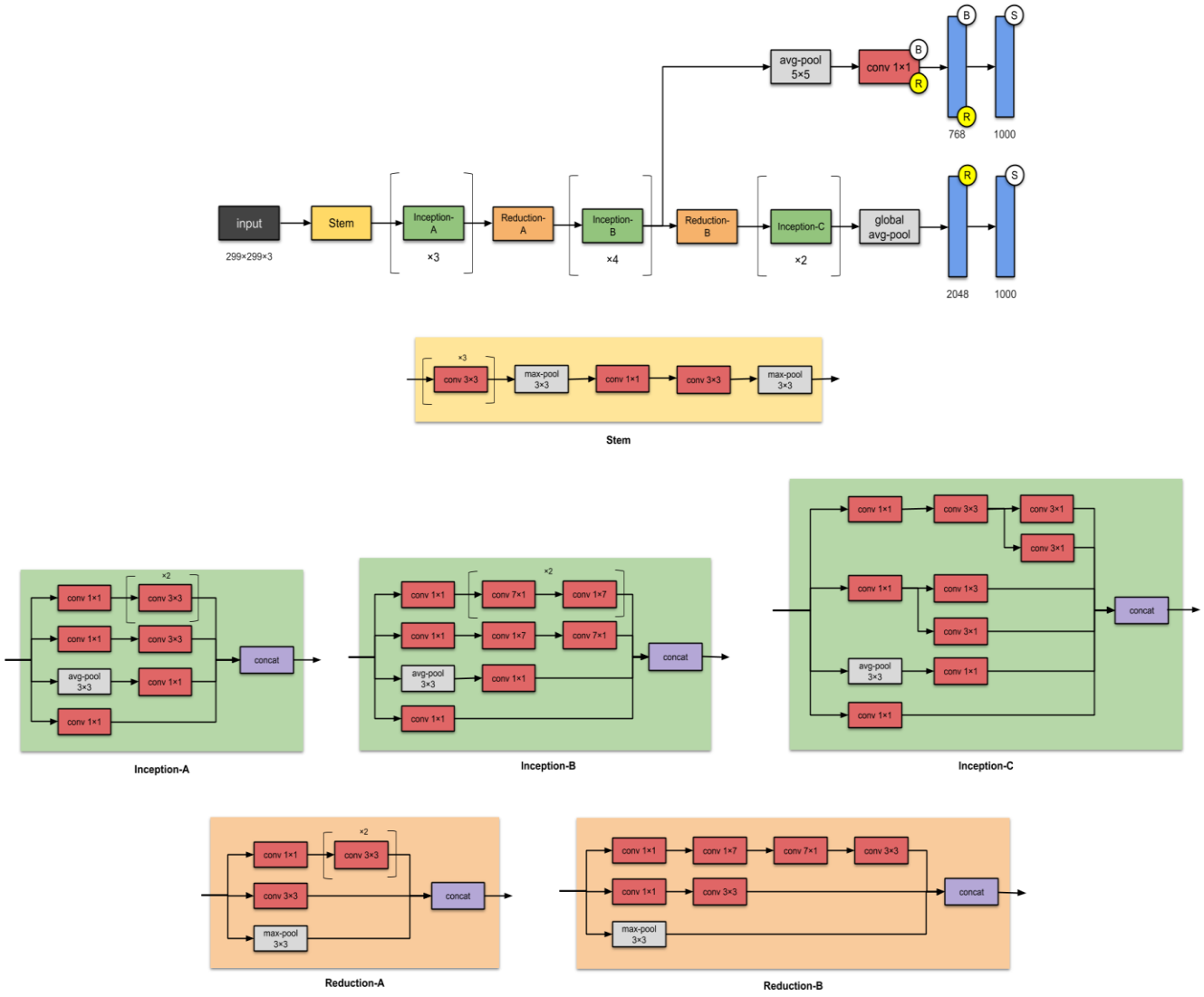


Figure A.3. Inception-v3 architecture[11].

[11]Szegedy, Christian, et al. "Rethinking the Inception Architecture for Computer Vision." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, doi:10.1109/cvpr.2016.308.

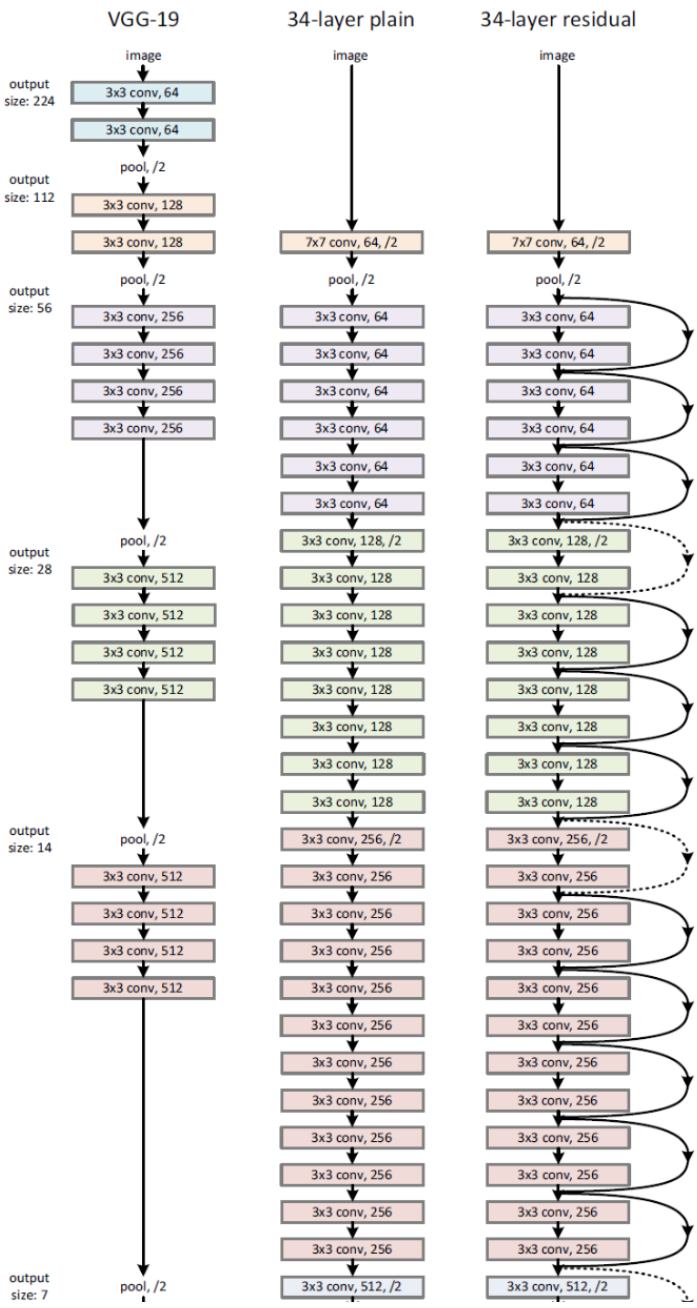


Figure A.4.ResNet Architecture.

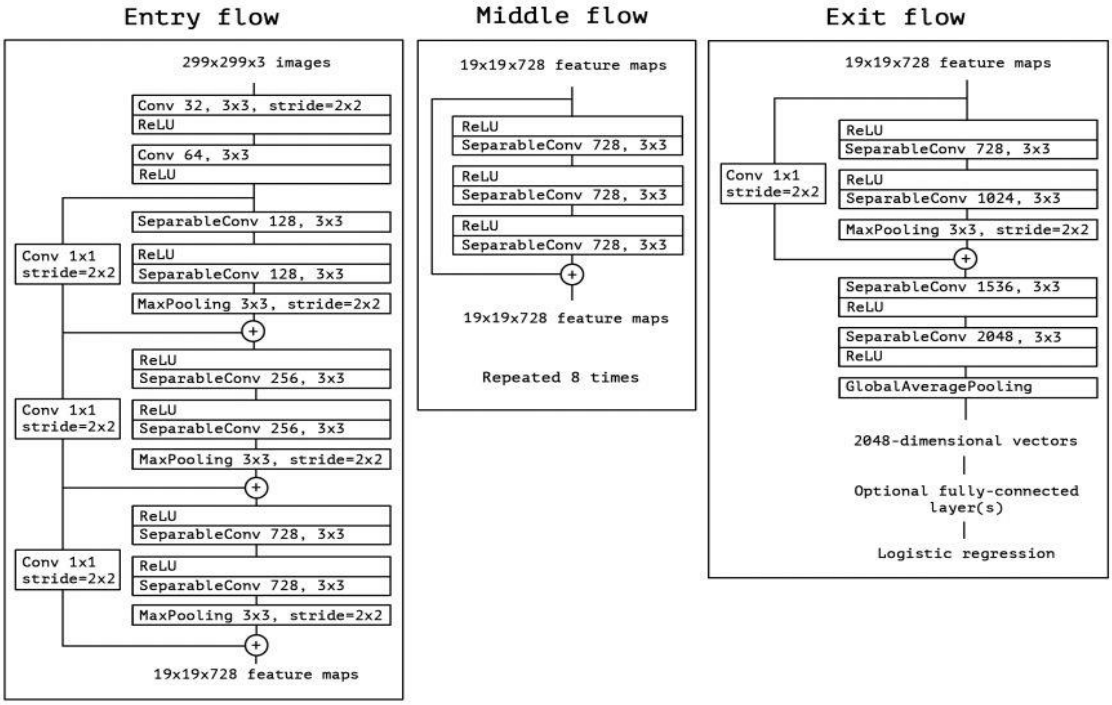


Figure A.5. Xception Architecture[12].

[12]Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Figure A.6.MobileNetV2 blocks Architecture.

Each line shows a sequence (blocks) of 1 or more identical layers that repeated (*n)* times, all spatial convolutions use 3 × 3 kernels, the stride (*s)* vary in each sequence use a 1 or 2. All layers in each sequence have the same number of output channel (*c).*