Democratic and Popular Algerian Republic

Ministry of Higher Education and Scientific Research

Yahia fares university of Médéa



Faculty of Sciences

Department of Mathematics and Computer Science

Final Master Project

Branch: computer science

Specialty: Systems Engineering and web Technologies

Theme

Colorectal abnormalities detection and classification using Deep Learning techniques

Prepared and presented by

BAKIR TERKI Othmane

Proposed and directed by

Dr.Sofiane TCHOKETCH KEBIR

Year: 2023-2024

Dedication

I dedicate this work to :

My dear parents,
My brothers and sisters,
My wife,
My angels: Mohamed Cherif, Tasnim and Kaouther.

Acknowledgment

Firstly, I would like to thank Almighty God, the most powerful and merciful, who gave me the strength and patience to complete this modest work. I would like to extend my gratitude to all my coworkers at the University Yahia Fares of Medea, who have provided for me throughout our studies. Through this modest work, I express my deep gratitude to my supervisor, Mr. TCHOKETCH KEBIR Sofiane, for accepting to guide me, providing me with interesting documentation, valuable advice, and all the facilities and advice he provided me during my study and completion of this project. I also want to express my sincere gratitude to the president and members of the jury for accepting to examine and evaluate my work. This work is the result of a laborious course of study where I was trained, followed, and supervised. Therefore, my deep gratitude and thanks are addressed to all of my teachers and supervisors, the entire teaching staff, administrators, and all school personnel.

Othmane

Contents

Dedication i

Acknowledgment ii

Table of contents iii

Figures list vi

Tables list viii

Abbreviations list ix

General Introduction 1

1 Medical imaging systems and their challenges 3

1.1 Introduction . . . 4

1.1.1 Medical imaging techniques . . . 4

1.1.2 Use of magnetic fields . . . 4

1.1.3 Using radioactivity : . . . 5

1.1.4 Use of X-rays : . . . 5

1.1.5 Use of ultrasound : . . . 6

1.1.6 Using light rays : . . . 6

1.1.7 Traditional OCT: . . . 6

1.2 Method of using colonscopy: . . . 6

1.2.1 Instrument: . . . 6

1.2.2 Visualization: . . . 6

1.2.3 Depth of Examination: . . . 7

1.2.4 Preparation: . . . 7

1.2.5 Sedation: . . . 7

1.2.6 Biopsy and Removal: . . . 7

1.2.7 Screening and Diagnosis: . . . 7

1.3 What is a colorectal cancer ? . . . 7

1.3.1 Overview: . . . 7
1.3.2 Colorectal Cancer Screening and Detection . . . 8
1.4 Clinical Background . . . 10
1.4.1 Colorectal polyp . . . 10
1.4.2 Target and objective . . . 11
1.5 Motivation . . . 11
1.6 Conclusion . . . 11

2 Deep Learning **13**
2.1 Introduction . . . 14
2.2 Machine Learning Background . . . 14
2.2.1 Deep Learning . . . 15
2.3 Artificial Neural Networks . . . 16
2.3.1 The Artificial Neural Network layers . . . 18
2.3.2 Activation Function . . . 19
2.3.3 Loss function: . . . 22
2.4 Convolutional Neural Networks . . . 23
2.4.1 Hyperparameters . . . 25
2.4.2 Autoencoder . . . 26
2.4.3 Approaches to Training . . . 27
2.5 Regularization Techniques: . . . 28
2.6 An overview of polyp classification and segmentation techniques . . . 29
2.7 Conclusion . . . 32

3 Colorectal Polyp Detection and Segmentation, DL implementation, Results and discussions **33**
3.1 Introduction . . . 34
3.2 Experiments Configuration . . . 34
3.3 Programming Language and Libraries . . . 34
3.3.1 Keras . . . 34
3.3.2 TensorFlow . . . 35
3.3.3 Python for Data Science . . . 35
3.4 DL Implementation, Experimental Results and discussion . . . 35
3.4.1 The used Dataset . . . 36
3.4.2 Data Preprocessing . . . 37
3.4.2.1 Data Augmentation . . . 37
3.4.2.2 Data splitting . . . 38
3.4.2.3 Data resizing . . . 38
3.4.3 The Proposed CNN deep Learning model . . . 39
3.4.4 Training and testing process . . . 43
3.5 Obtained results and discussion . . . 43

3.5.1 Evaluation metrics . . . 43
3.5.1.1 Dice Coefficient and Jaccard: . . . 43
3.5.1.2 Recall . . . 44
3.5.1.3 Precision . . . 44
3.5.1.4 Accuracy . . . 44
3.5.1.5 F1 score . . . 44
3.5.1.6 ROC curve . . . 45
3.5.1.7 Specificity . . . 45
3.5.2 Obtained results and discussion . . . 45
3.5.2.1 Global obtained results . . . 45
3.5.2.2 Best and Worst obtained results: . . . 48
3.6 Conclusion . . . 50

General Conclusion **51**

Abstract **56**

Figures list

1.1 Endoscopic appearance of colorectal cancer:(A) Hyperplastic polyp (B) Sessile Serrated Adenoma. . . . 8
1.2 Images taken by a colonoscope. . . . 9
1.3 Wireless capsule endoscopy. . . . 9
1.4 Images taken by a wireless capsule. . . . 10
1.5 Parts of the Colon with polyp. . . . 10

2.1 The Biological neuron. . . . 16
2.2 Small network of neurons: the dendrites receive the input signals, the body of the cell (nucleus) is responsible for processing the input signals and the axon sends the signal from the body of the cell towards the neighbor neurons. 16
2.3 Artificial neural network. . . . 17
2.4 Perceptron Model. . . . 18
2.5 Artificial Neural Network architecture. . . . 18
2.6 activation function. . . . 19
2.7 Relu function. . . . 20
2.8 Tanh function. . . . 21
2.9 Softmax function. . . . 21
2.10 Loss function. . . . 22
2.11 Prototype deep CNN with 2 hidden layers, 2 pooling layers, fully linked layer, and 9 output units. . . . 23
2.12 Autoencoder architecture . . . 27
2.13 Autoencoder schema . . . 27

3.1 A representation of the Deep Learning model for colorectal polyp detection and classification. . . . 36
3.2 Sample images of CVC-ClinicDB. Original images shown in the first row and their corresponding ground truth shown in the second row. . . . 37
3.3 Sample augmented polyp images with different shapes and sizes. . . . 38
3.4 U-net synoptic model. . . . 42

3.5 Polyp colorectal detection and of segmented (images index 1) 45
3.6 Polyp colorectal detection and of segmented (images index 320) 46
3.7 Polyp colorectal detection and of segmented (images index 470) 46
3.8 Polyp colorectal detection and of segmented (images index 612) 47
3.9 Best obtained results for image 306 49
3.10 Worst obtained results for image 155 49

List of Tables

2.1 Summary of approaches proposed for colorectal cancer classification. . . . 29

3.1 Training-testing split details 38
3.2 The calculation of TP, TN, FN, and FP. 44
3.3 Metric Values (Mean ± Std Dev) 47
3.4 Summary of Dice Coefficient Metrics 48
3.5 Summary of Jaccard Index Metrics 48

List of Acronyms and Glossary

MRI: Magnetic resonance imaging.

ADR: Adenoma Detection Rate.

CNN: Convolutional Neural Network.

AI: Artificial Intelligence.

CAD: Computer Aid Diagnosis.

CPU: Central Processing Unit.

CRC: Colorectal Cancer.

DL: Deep Learning.

DCNN: Deep Convolutional Neural Network.

FCN: Fully Convolutional Networks.

FC: Fully Connected.

ML: Machine Learning.

SSA: Sessile Serrated Adenoma.

SVM: Support Vector Machine.

ReLU: Rectified Linear Unit.

VGG: Visual Geometry Group.

WCE: Wireless Capsule Endoscopy.

SPECT: Single-Photon Emission Computed Tomography.

PET: Positron Emission Tomography.

OCT: Optical Coherent Tomography.

ANN:: Artificial Neural Networks.

TP: True Positive.

FP: False Positive.

TN: True Negative.

FN: False Negative.

ROC: Receiver Operating Characteristics.

GPU: Graphics Processing Unit.

RAM: Random Access Memory.

General introduction

Over the past two decades, the rapid development of precision investigation systems in the field of medical imaging has opened up new horizons in medical horizons in medical analysis, interpretation and treatment.

The integration of these systems with new high-precision technologies has made it possible to achieve quality diagnoses and much faster, more precise surgical interventions. These new systems include computer-assisted surgery systems, medical artificial intelligence systems, medical imaging systems, etc.

Since they were first discovered, medical imaging systems have remained the most effective means of investigating and accurately analyzing a wide range of pathologies. most effective means of investigating and accurately analyzing a wide variety of patients.

In fact, they are the only means of visualizing in vivo the anatomy (structure) and function of the human body by photographing it without the need for surgery. It also provides essential and important in the form of 2D and 3D medical images about the human body, organs and even human cells.

At present, most medical imaging systems produce a significant amount of medically relevant data that is still hidden at the pixel intensity level. As a corollary, the integration of algorithmic tools and methods is required to exploit and extract this medical information hidden in the data to obtain precise and effective medical interpretations and interventions.

According to statistics from the World Health Organization, colorectal cancer ranks as the third most common type of cancer in terms of incidence and mortality rates globally.

Optical colonoscopy is a crucial diagnostic and therapeutic approach for colorectal cancer. It allows physicians to visually examine the colon tissue in a living patient and also facilitates the removal of visible colorectal lesions or abnormalities

The remarkable progress made in medical image analysis underlines the importance of integration of these algorithmic tools and methods as accurate, fast in clinical and medical settings. Nevertheless, new challenges have emerged in the analysis and interpretation of medical images, such as :

1. Need for new techniques and algorithmic methods for image analysis that are

more efficient and can be adapted to a specific clinical task (anomaly detection anomalies, etc.).

2. Need for reliable methods and techniques to produce large-scale machine learning databases that meet the growing demands of researchers for a successful to successfully complete the learning phase in supervised methods and techniques based on Deep learning and machine learning techniques. learning techniques.

3. Need for methods and algorithms to analyze heterogeneous image data with minimum human interaction.

4. Need for more compliant image analysis methods to build patient-specific models (the manufacture of prostheses and the implantation of vital parts or whole organs).

In this dissertation, we present methods and algorithmic techniques adapted to medical image analysis for detecting and segmenting colorectal abnormalities, specifically colon polyps.

Chapter 1

Medical imaging systems and their challenges

1.1 Introduction

Over the past two decades, the rapid development of precision investigation systems in the field of medical imaging has opened up new horizons in medical analysis, interpretation, and treatment. The integration of these systems with new high-precision technologies has made it possible to achieve quality diagnoses and much faster, more precise surgical interventions. Since they were first discovered, medical imaging systems have remained the most effective means of investigating and accurately analyzing a wide range of pathologies.

1.1.1 Medical imaging techniques

Depending on the techniques used, medical imaging examinations can provide information on organ anatomy (size, volume, location, shape of any lesions, etc.) or function (physiology, metabolism, etc.). The former is known as structural imaging, the latter as functional imaging.

The structural imaging methods most commonly used in medicine include tomographic methods based either on X-rays (conventional radiology, computed tomography or CT-scan, angiography, etc.) or magnetic resonance (MRI), ultrasound methods (using ultrasound), and optical methods (using light rays), colonscopy.

Functional imaging methods are also highly varied. They include nuclear medicine techniques (PET, SPECT) based on the emission of gamma rays by radioactive tracers which, after injection, concentrate in regions of intense metabolic activity, electrophysiological techniques which measure changes in the electrochemical state of tissues (particularly in relation to nerve activity), and thermographic measurements.

These different types of techniques are often used in a complementary way, sometimes even within the same imaging system, enabling multimodal acquisitions, whether simultaneous or not [1].

1.1.2 Use of magnetic fields

Magnetic resonance imaging (MRI) uses the effect of an intense magnetic field on proton spin. It is a tomographic process, providing "virtual slices" of the body in three spatial planes (sagittal, coronal and axial).

Depending on the parameters chosen, MRI can provide highly contrasted images of certain tissues according to their histological properties. It is therefore a particularly useful tool for brain imaging. To date, MRI examinations are considered risk-free for the body. However, all ferromagnetic objects sensitive to magnetic fields (piercings, pacemakers, certain prostheses, etc.) are dangerous.[2]

Magnetoencephalography (MEG) is a technique for measuring weak magnetic fields induced by the electrical activity of brain neurons. Unlike MRI, it does not rely on the

prior magnetization of tissue. Consequently, the presence of magnetic objects poses no risk.

Magnetocardiography is a technique very similar to the previous one, involving the measurement of magnetic fields induced by the electrical activity of heart muscle cells in the torso. It is rarely used.[1]

1.1.3 Using radioactivity :

Nuclear scintigraphy techniques are based on the use of a radioactive tracer that emits radiation that can be detected by measuring equipment. These radiopharmaceutical molecules are chosen to bind preferentially to certain cells depending on the type of diagnosis required.

Computer processing of the data then enables the spatial origin of this radiation to be reconstructed and the regions of the body where the tracer has concentrated to be deduced.

The image obtained is usually a projection, but it is also possible to obtain a cross-section or 3D reconstruction of the distribution of the tracer.

Single-photon emission computed tomography (SPECT): this uses gamma photon emission from a molecule labelled with a radioactive isotope injected into the body.

Positron emission tomography (PET): this usually uses sugar (a glucose analogue) labelled with a radioactive substance that emits positrons (e.g. Fluor 18), making it possible to see cells with a high metabolic rate (e.g. cancer cells, infection, etc.).

PET generally provides better quality images than SPECT. However, the number and availability of radiopharmaceuticals that can be used in SPECT and the low cost of gamma cameras compensate for this shortcoming.[1]

1.1.4 Use of X-rays :

X-rays are commonly used. These rays, like gamma rays, are ionising and therefore dangerous. In particular, irradiation of a cell undergoing mitosis can cause DNA mutation, which can ultimately lead to cancer. However, thanks to radiation protection measures, the risk inherent in X-ray examinations is limited as much as possible.[1]

Different types of examination use X-rays :

Radiography, using X-rays and sometimes the injection of a contrast product. The images obtained are projections of the organs and various systems in a plane. Radiography is generally used for the bone system, as this is the most visible system on a body X-ray.

X-ray scanner, tomography using X-rays. The images obtained are millimetre (or infra-millimetre) slices that can be studied in all planes of space, as well as three-dimensional images. DEXA scanner measuring bone density (or osteodensitometry).[1]

1.1.5 Use of ultrasound :

Ultrasound, using ultrasound. The image obtained is a cross-section of the organ studied. It can be combined with a Doppler examination analysing the speed of blood in the vessels or heart cavities or with a measurement of Young's modulus by coupling to a low-frequency vibration (technique developed in 2005).

1.1.6 Using light rays :

OCT (Optical Coherent Tomography) technologies make it possible to obtain an image by producing optical interference beneath the surface of the tissue being analysed. These interferences are measured by a camera (full-field OCT) or by a dedicated receiver (traditional OCT). These techniques are non-destructive and safe.

Full-field OCT is the most powerful of all OCT techniques. The image obtained is a virtual optical biopsy. This technology is currently being developed and, thanks to its resolution (1 micro meter in the 3 dimensions X, Y, Z), it allows the cellular organisation to be seen in 3 dimensions. The images are taken in plan, like photos taken above the tissue, but at different depths below the surface of the tissue being observed. This technique uses a white light source (broad spectrum).[1]

1.1.7 Traditional OCT:

the image obtained is a section of the tissue studied. Resolution is of the order of 10 to 15 micro meter. This technology uses a laser to produce the images.

1.2 Method of using colonscopy:

Colonoscopy is a medical procedure that allows for the visual examination of the interior lining of the colon (large intestine) and the rectum. It is performed using a flexible, slender instrument called a colonoscope, which is equipped with a small camera and a light source. Some key parameters and characteristics of colonoscopy include:

1.2.1 Instrument:

The colonoscope is a long, flexible tube with a camera and light source at the tip. It is typically between 4 and 6 feet long and has a diameter of around 0.5 to 0.8 inches.

1.2.2 Visualization:

The camera at the end of the colonoscope transmits live video images from inside the colon to a monitor, allowing the doctor to examine the lining of the colon for abnormalities such as polyps, tumors, or inflammation.

1.2.3 Depth of Examination:

A complete colonoscopy allows for the visualization of the entire colon, from the rectum to the cecum (the beginning of the colon, where it joins the small intestine).

1.2.4 Preparation:

Patients are typically required to follow a clear liquid diet and take laxatives or enemas to empty their colon before the procedure.

1.2.5 Sedation:

Colonoscopy is usually performed under sedation, with the patient receiving medications to help them relax and reduce any discomfort during the procedure.

1.2.6 Biopsy and Removal:

The colonoscope can be equipped with instruments that allow the doctor to take tissue samples (biopsies) or remove polyps or other abnormal growths during the procedure.

1.2.7 Screening and Diagnosis:

Colonoscopy is a crucial tool for screening and diagnosing conditions such as colorectal cancer, inflammatory bowel diseases (e.g., Crohn's disease, ulcerative colitis), and other disorders affecting the colon.

These parameters and characteristics make colonoscopy a valuable diagnostic and therapeutic tool in the field of gastroenterology and colon health management.

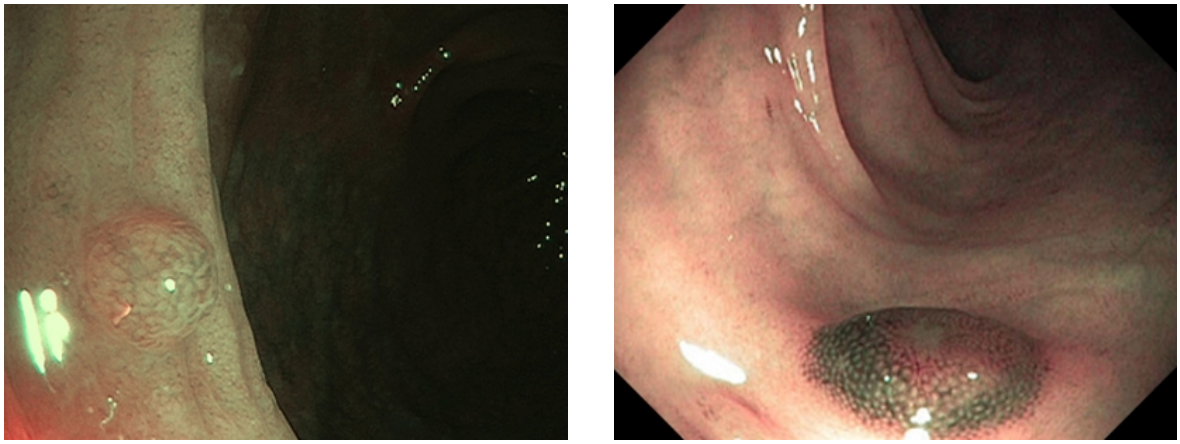
1.3 What is a colorectal cancer ?

1.3.1 Overview:

colorectal cancer is the second most common cause of cancer-related death worldwide among the various gastrointestinal diseases, which can range from minor irritations to fatal conditions.[3] Adenomas of the colorectal cavity Polyps are tiny cell tumors that develop in the colon's lining .[4] While the majority of colon polyps are harmless and typically do not produce symptoms, some may eventually develop into cancer. [5] Colorectal adenomas progress slowly to become dangerous adenocarcinomas. Consequently, the prevention of colorectal cancer and the safe treatment of it depend heavily on early identification, endoscopic evaluation, and histological characterization.[6] One of the most important strategies for finding polyps and stopping their growth is colorectal cancer screening.[7]Stool DNA testing, wireless capsule

endoscopy, computed tomography colonography, and colonoscopy are a few common colorectal cancer screening methods. Colonoscopy is currently the most widely utilized screening procedure for colorectal polyps.[8] The difficult aspect of these tests, however, is the follow-up process following screening or CRC surveillance, in which no polyp is found but some that are may progress to a different stage before the subsequent examination.

Numerous research in the field of clouding [9] have demonstrated that colonoscopists frequently miss tiny and occasionally larger colorectal polyps. Furthermore, it can be difficult to accurately characterize the type of polyp. For example, it can be difficult to distinguish between sessile serrated polyp and hyperplastic, as seen in Figure 1.1, which is not a good fit for any screening technique [10].



A) Hyperplastic polyp B) Sessile Serrated Adenoma

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

1.3.2 Colorectal Cancer Screening and Detection

Screening of colon cancer can detect early tumors and polyps in the big gut. This form of screening may be able to discover the abnormalities that can be treated prior to growing or propagating cancer. Regular examinations can lower death risk and colorectal cancer complications. Several test options are available for colorectal cancer screening such as the following:

Colonoscopy is a laparoscopic examining 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 within the body, and attached directly to the TV screen allows the doctor to watch the whole colon, If necessary, polyps or other abnormal tissue can be removed

using the endoscope during the colonoscopy procedure, or to halt bleeding, as well as tissue samples (biopsies) can also be obtained during colonoscopy.[11] An image taken through colonoscopy is displayed in Figure2.2.

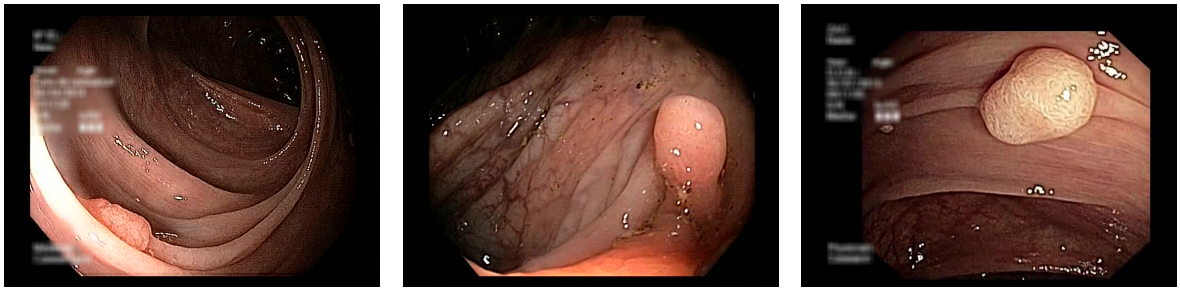


Figure 1.2. Images taken by a colonoscope.

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.[11]

Wireless Capsule Endoscopy Wireless Capsule Endoscopy a medical treatment used to diagnose and visualize the gastrointestinal tract. The endoscopy is carried out utilizing a small customized wireless camera put within a capsule swallowed by the patient, shown in Figure2.3.The capsule is compact in size, enabling it to reach locations that the regular endoscope cannot reach, as well as the sharpness of its images and the magnification it gives.[11]



Figure 1.3. Wireless capsule endoscopy.

Thousands of images of the stomach, intestines, and other organs are captured by the capsule as it passes through the digestive system; sample images from a WCE are displayed in Figure 2.4. It is captured and transmitted to a fixed receptor worn around the patient's waist. The capsule exits the patient's body along with their stool after roughly eight hours. After that, it is not utilized again.

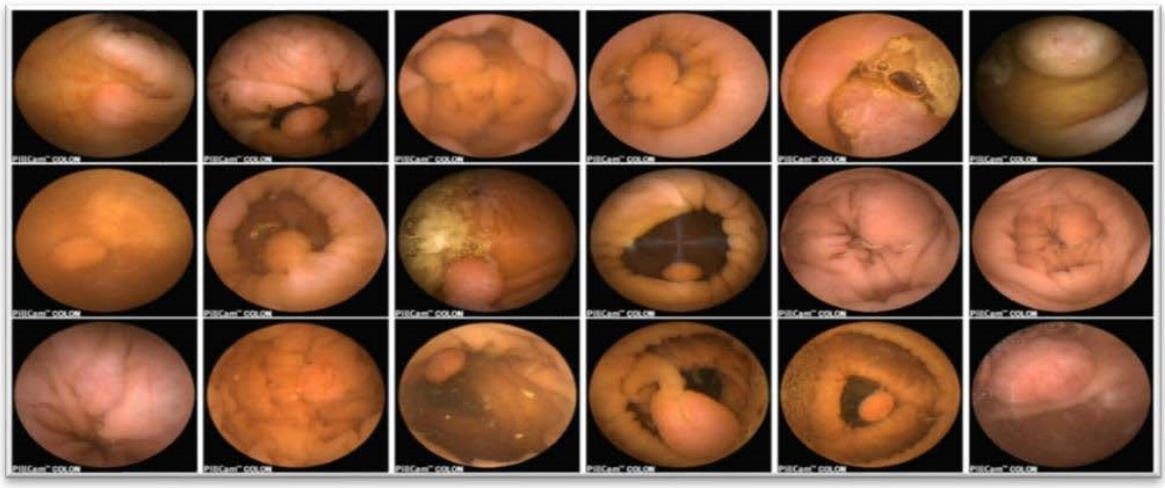


Figure 1.4. Images taken by a wireless capsule.

The ensuing parts will delve into a detailed discussion of machine learning, encompassing various types. They will also show the processes of artificial neural networks and convolutional neural networks, and ultimately provide illustrations of the suggested approaches for colorectal polyps and their applications.[11]

1.4 Clinical Background

This section provides an explanation of polyp and process of growths, given different modalities for screening testing.

1.4.1 Colorectal polyp

Polyps are abnormal growths of tissue from a mucous membrane (inner lining) of the colon or rectum. As indicated in Figure 2.1, polyps typically appear in many regions of the gastrointestinal system, such as the stomach, ear, nose, urinary bladder, colon, and uterine. Some polyps are flat, it is said to be sessile whereas others attached to the surface by a narrow elongated stalk [4].

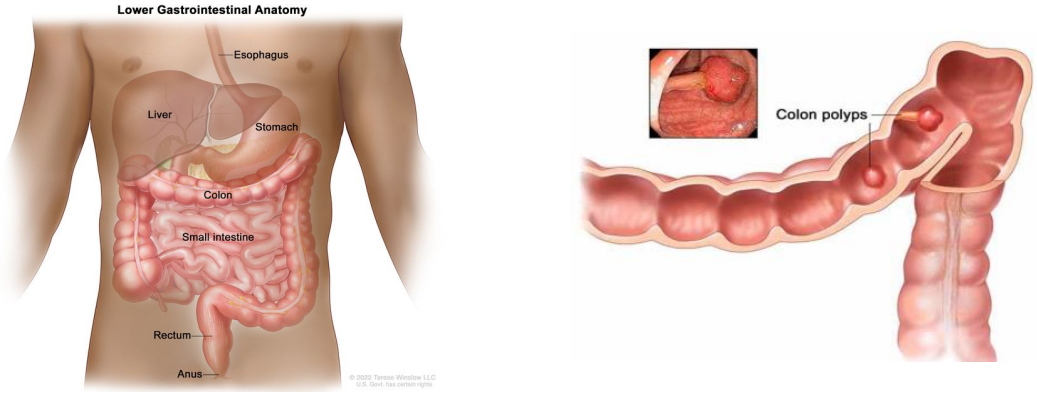


Figure 1.5. Parts of the Colon with polyp.

Pathologists face a difficult problem when deciding whether to remove or ignore polyps since it requires skills that many endoscopists lack [12]. The reason is that in order to perform a biopsy, patients must first have a colonoscopy examination. Once a polyp is detected, small samples of its suspicious sections must be removed; these samples should subsequently be analyzed by a histopathologist. It is evident that this procedure causes inconvenience and takes a long time for practitioners to identify the type of polyp.[13]

Physicians and computer vision researchers must move quickly to address the issues and develop more precise and dependable techniques for polyp detection and classification based on endoscopic images in order to address the early detection and prevention of colon cancer. As one of the key components used in numerous intelligent information technologies, machine learning has advanced considerably [14].

1.4.2 Target and objective

The aim of this work is to use image segmentation techniques, specifically deep learning-based methods, to accurately detect and classify regions containing well-defined polyps in medical images. By precisely segmenting and delineating polyp boundaries, the research enables characterization of polyp morphology and classification into clinically relevant categories. This comprehensive analysis of segmented polyps facilitates determining the ideal therapeutic strategy, such as polyp removal or continued surveillance. Ultimately, the segmentation-driven polyp detection and classification framework aims to enhance early diagnosis of colorectal abnormalities, reduce missed polyps, and improve the efficacy of colorectal cancer screening and prevention efforts.

1.5 Motivation

Recent advances in research expounded that deep learning networks are particularly influential for autonomous image processing. Deep learning models have shown huge increases in the performance of several medical applications. This possess our curiosity to develop and conduct an effective approach based on Deep Learning algorithms to detect the discriminative characteristics in colorectal polyp images, which in turn will lead to an effective classification and Detection model that can potentially be a life savior.

1.6 Conclusion

This work aims to leverage advanced deep learning techniques for the accurate detection and classification of colorectal polyps from medical images. By employing precise segmentation algorithms, determining polyp regions and describing their morphology is the objective of the research in order to enable accurate classification into clinically relevant classes. The main objective is to enable early diagnosis, reduce missed

polyps and improve the results of colorectal cancer screening and prevention initiatives. Recent developments in deep learning results in excellent results in a range of medical applications, motivating the developing of an effective approach dedicated especially for colorectal polyp detection. By enabling the best possible interventions and medical decision-making, the proposed segmentation-driven polyp detection and classification system has the potential for greatly improving the results for patients and colon health care in general. Ultimately, this research contributes to the broader field of medical imaging and computer-assisted diagnosis, addressing a crucial healthcare challenge with profound implications for colorectal cancer prevention and treatment.

Chapter 2

Deep Learning

2.1 Introduction

This chapter provides a comprehensive overview of deep learning techniques and their application in the context of colorectal polyp detection and classification. It begins by establishing the background of machine learning, exploring its various types, including supervised, unsupervised, semi-supervised, and reinforcement learning, and highlighting their relevance in the medical field, particularly in colorectal cancer research. The chapter then delves into the fundamentals of artificial neural networks (ANNs), detailing their structure, layers (input, hidden, output), and the role of weights and biases in the learning process. It also elaborates on activation functions, such as sigmoid, ReLU, tanh, and softmax, and their significance in introducing non-linearity and enabling complex pattern recognition in ANNs. Additionally, the chapter discusses loss functions, which are crucial for evaluating model performance and guiding the training process.

2.2 Machine Learning Background

Artificial intelligence has recently captivated the world's interest in numerous forms [15]. Several models demonstrate that artificial intelligence has approached competition from human intellect, with this success, ever more attention with what is known as machine learning to go towards greater gaining in utilizing artificial intelligence. ML is one of AI areas based on constructing computer programs in many forms, which increase their performance at some task via the experience. The term machine learning was developed by the pioneer of artificial intelligence Arthur Samuel in 1959 while working in IBM laboratories [16], he utilized 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 developed and made open to permit more folks to manufacture and train AI models [17].

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 uncover the advanced principles. ML techniques are fundamentally three: supervised, reinforced, and unsupervised. These approaches are addressed briefly in the following subsections:

• Supervised Learning :

is a machine learning method aimed at developing a model capable of predicting or classifying unknown data from a set of labeled examples and associating the inputs with the desired outputs. The process is divided into two main phases: the learning phase and the testing phase. During the training phase, the algorithm is trained with a set of labeled data, where each example is associated with a specific label. The testing phase then involves using this model to predict the labels of new, previously unseen data, thereby evaluating the generalization ability of the algorithm. Learning effectiveness is measured by the accuracy of the algorithm's predictions compared to the actual labels provided by

the experts.[18]

• Unsupervised Learning :

This machine learning modality is deployed in the absence of supervision by experts. It focuses on the structuring and categorization of information without prior knowledge of the labels or classes assigned to the data. The main objective is the identification of hidden structures or unknown patterns within the data, thus facilitating the understanding and exploitation of unlabeled information.[18]

• Semi-supervised Learning :

Semi-supervised learning is a hybrid learning approach that uses both labeled and unlabeled data to train machine learning models. In many real-world scenarios, obtaining a large set of labeled data can be costly and labor-intensive. However, unlabeled data is often abundant and cheaper to collect. Semi-supervised learning combines a small amount of labeled data with a large amount of unlabeled data during the training process. Semi supervised is applied for metastatic colon lymph node diagnosis [19].

• Reinforcement learning :

This learning approach is characterized by the interaction of an algorithmic agent with its environment, with the aim of optimizing a reward function through the selection of sequences of actions. The agent learns to execute actions that maximize cumulative rewards, based on the feedback (rewards or penalties) received from the environment following its actions. This method finds significant applications in various fields such as robotics, game theory, and autonomous vehicle navigation, where it allows the agent to adapt its strategies autonomously to achieve its goals[18].

2.2.1 Deep Learning

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics.

Deep learning discovers intricate structure in large data sets by using the back-propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep Convolutional networks have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent networks have shone light on sequential data such as text and speech[20].

2.3 Artificial Neural Networks

Artificial Neural Networks are a collection of mathematical models that are based on the structure of biological neural networks, where the neuron is the fundamental processing unit of the brain [21]. Neurons communicate with one other by aggregating impulses from linked neurons [21]. When the cumulative stimuli surpass a specific threshold, the neuron undergoes activation, resulting in the generation of a new stimulus that is transmitted across the network. The information is stored in the synaptic connections between the neurons in the brain[21].

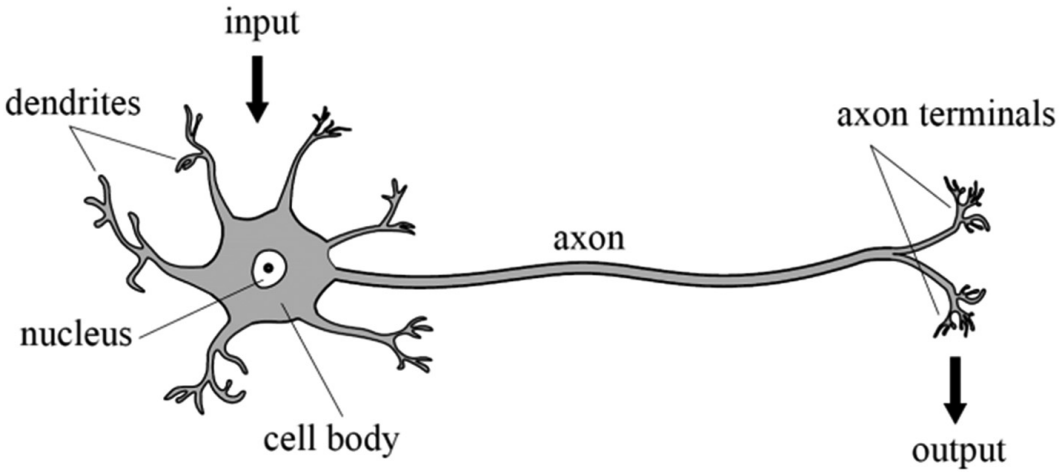


Figure 2.1. The Biological neuron.

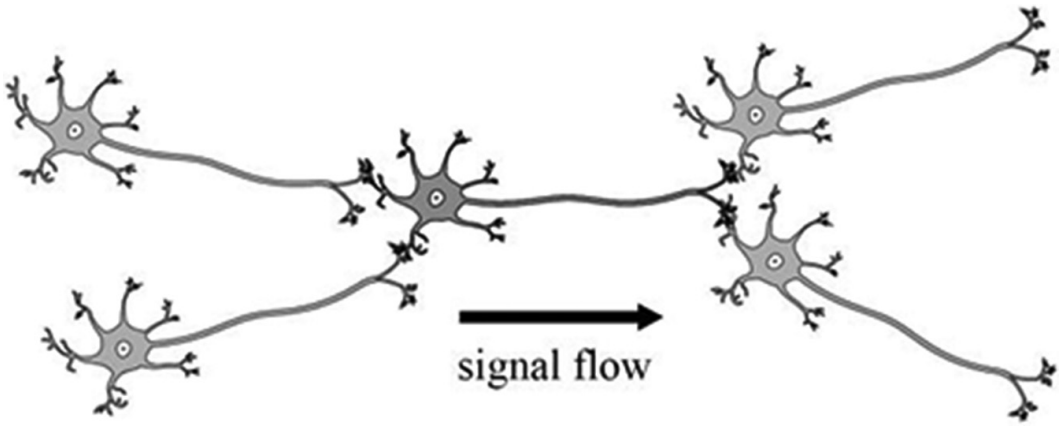


Figure 2.2. Small network of neurons: the dendrites receive the input signals, the body of the cell (nucleus) is responsible for processing the input signals and the axon sends the signal from the body of the cell towards the neighbor neurons.

Accordingly, artificial intelligence has tended to construct artificial neural networks that mirror the human brain. Artificial neural networks consist of neurons that communicate with each other via linkages. These linkages are utilized 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 by this link.

As illustrated in Figure 2.3, the data that need to be processed is placed at the first layer of units $(x_1, x_2, \ldots, x_i)$. Each input is connected to neurons (j) and weights $(w_{1j}, w_{2j}, \ldots, w_{ij})$.

Every signal is multiplied with its relevant weights on the connection, then the neuron sums all signals received and sends them to the non-linear activation function $\phi(\cdot)$ in order to generate the final output (y).

The input layer of an ANN consists of n input units with values x_i, where $i = 1, 2, \ldots, n$, and randomly determined initial weights, usually from the interval $[-1, 1]$. Each unit in the hidden (middle) layer receives the weighted sum of all x_i values as the input. The output of the hidden layer, denoted as Y_c, is computed by summing the inputs multiplied by their weights and adding a bias term b, according to Equation [2.1]:[22]

$$Y_c = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{2.1}$$

where f is the activation function selected by the user (sigmoid, tangent hyperbolic, exponential, linear, step or other).

the bias term b is typically represented by an additional term in the weighted sum calculation. This allows the neuron to learn a different offset for the activation function, which can be useful for modeling complex relationships between inputs and outputs.

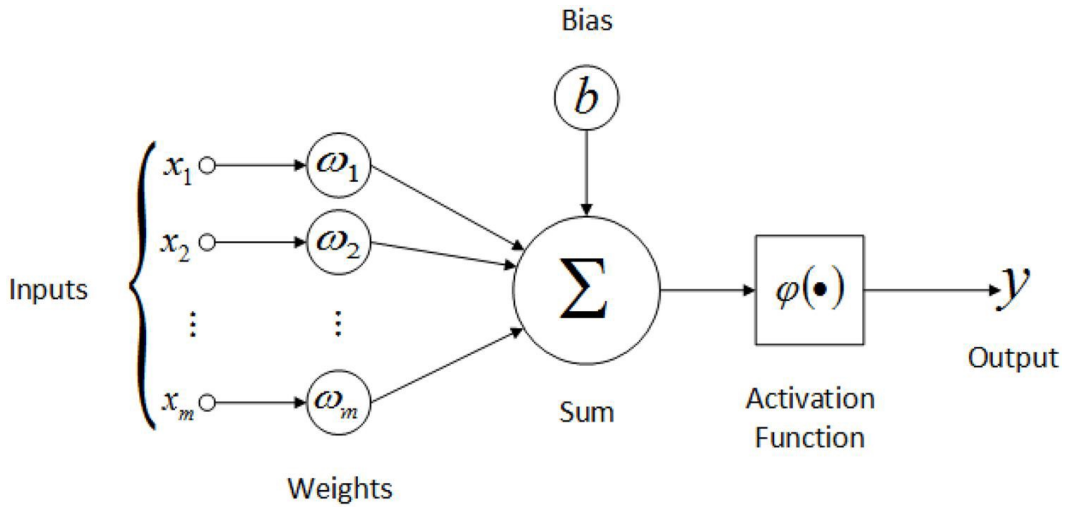


Figure 2.3. Artificial neural network.

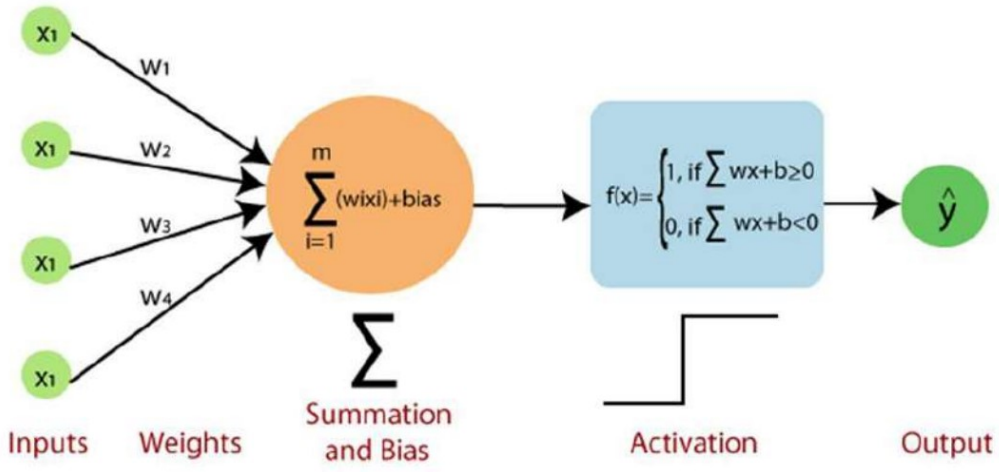


Figure 2.4. Perceptron Model.

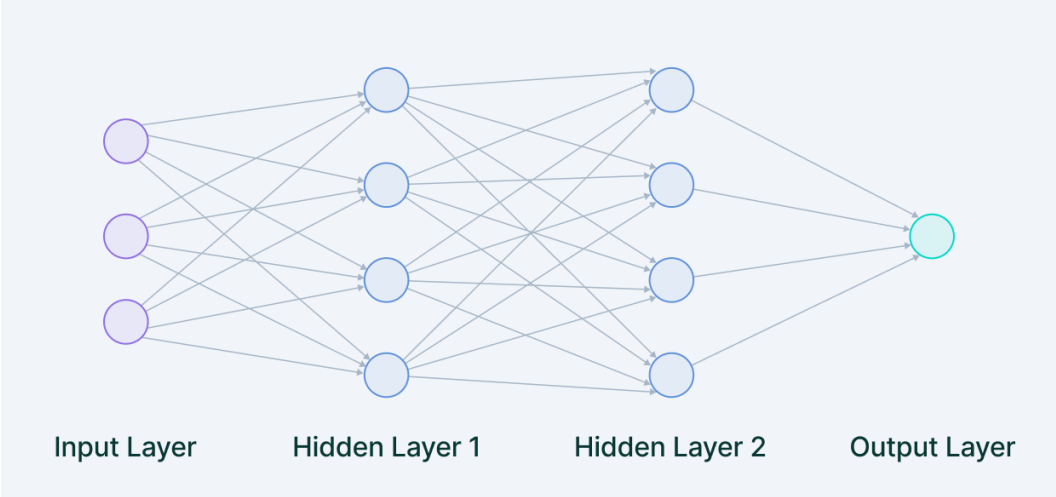


Figure 2.5. Artificial Neural Network architecture.

2.3.1 The Artificial Neural Network layers

• **Input layer :** It is the set of features that are fed into the model for the learning process. For example, the input in object detection can be an array of pixel values pertaining to an image.[23]

• **Hidden layer :** The second type of layer in a neural network is the hidden layer, these are the layers between the input and output layers, composed of neurons, each of them is connected to every neuron in the adjacent layers. These layers consist of several interconnected nodes that perform computations on the data received from the input layer.

• **Output layer:** It is the last layer of a neural network , it represents the final prediction generated by the network, This layer may consist of one or more nodes , at each node, an activation function is applied to transform the weighted sum of the inputs from the previous layer into an output depending on the problem being solved.

• **Weight:** Its main function is to give importance to those features that contribute more towards the learning. It does so by introducing scalar multiplication between the input value and the weight matrix. For example, a negative word would impact the decision of the sentiment analysis model more than a pair of neutral words.[23]

• **Bias:** The role of bias is to shift the value produced by the activation function. Its role is similar to the role of a constant in a linear function.[23]

2.3.2 Activation Function

It is also known as Transfer Function, is to combine multiple inputs into one output value so that the activation function can be applied. It is done by a simple summation of all the inputs to the transfer function, It introduces non-linearity in the working of perceptrons to consider varying linearity with the inputs. Without this, the output would just be a linear combination of input values and would not be able to introduce non-linearity in the network.[23]

Sigmoid function : The Sigmoid functions are simply mathematical equations that equate all values of x to an output ranging between 0 and 1 thus creating an S-shaped graph. This was a great step in neural networks as it allowed the outputs of these functions to be more in line with standard probability situations. The most common forms of sigmoid functions that are typically found in artificial neural networks are:

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{or} \quad \frac{e^x}{e^x+1} \tag{2.2}$$

The graphical representation of this activation function can be found in Fig. 2.6.

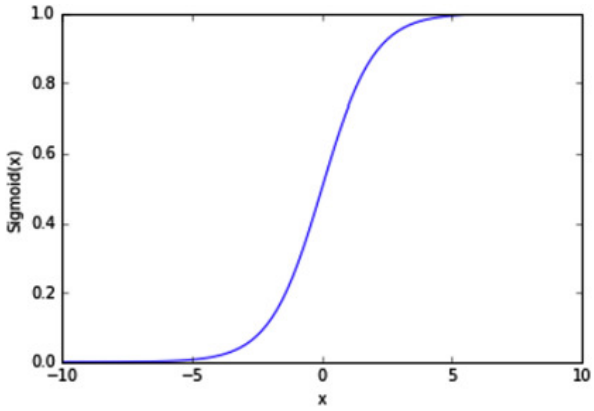


Figure 2.6. activation function.

Relu function : ReLU (Rectified Linear Unit) Activation Function is a mathematical function commonly used in artificial neural networks. It is applied to the output of a neuron to introduce non-linearity and enable the network to learn complex patterns and relationships in the data. ReLU function outputs the

input value if it is positive, and if the input value is negative, it outputs zero. It works by applying a simple mathematical operation to the input value. If the input value is greater than or equal to zero, the output is equal to the input value. If the input value is negative, the output is zero.[24] Mathematically, the ReLU function can be represented as:

$$f(x) = \max(0, x) \tag{2.3}$$

The graphical representation of this Relu function can be found in Fig. 2.7.

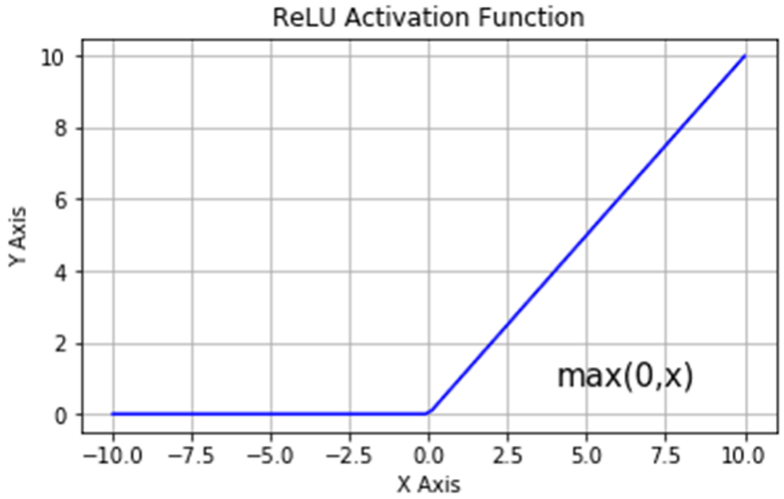


Figure 2.7. Relu function.

Tanh function : The tanh activation function, also called the hyperbolic tangent activation function, is a mathematical function commonly used in artificial neural networks for their hidden layers. It transforms input values to produce output values between -1 and 1. It is expressed as the ratio of the difference between the exponential of the input value and the exponential of its negation to the sum of these exponentials.[25] Mathematically, the tanh activation function can be represented as:

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.4}$$

The graphical representation of this Tanh function can be found in Fig. 2.8.

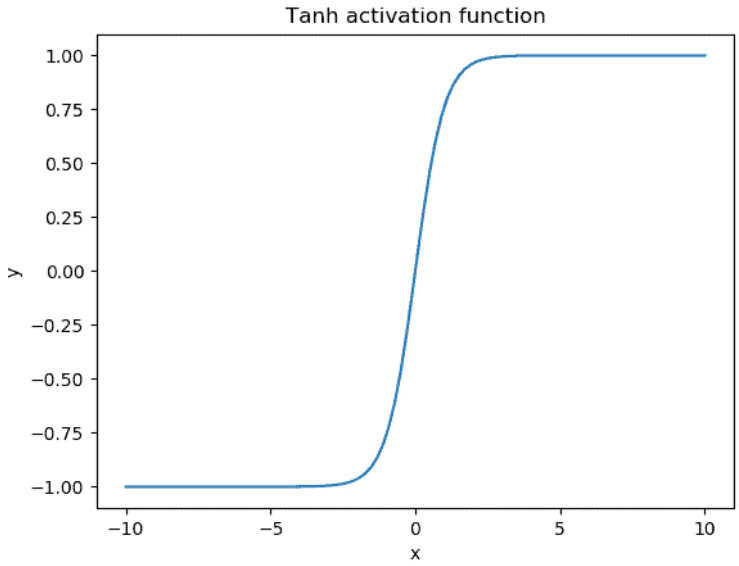


Figure 2.8. Tanh function.

Softmax function : The softmax function is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1.[26]Mathematically, the Softmax function can be represented as:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{2.5}$$

The graphical representation of this Softmax function can be found in Fig. 2.9.

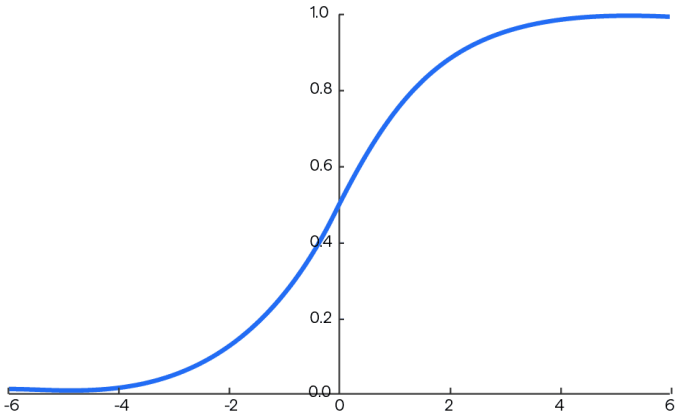


Figure 2.9. Softmax function.

2.3.3 Loss function:

The loss function is the function that computes the distance between the current output of the algorithm and the expected output. It's a method to evaluate how your algorithm models the data. It can be categorized into two groups. One for classification (discrete values, 0,1,2...) and the other for regression (continuous values).[27] This loss function measures the performance of a classification model whose output is a probability value between 0 and 1. It quantifies the difference between the predicted probabilities and the actual class labels. The goal of the model is to minimize this loss during training. Mathematically, the Loss function can be represented as:

$$\text{LCE} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{2.6}$$

Where:

- L_{CE} is the cross-entropy loss.
- n is the number of samples.
- y_i is the true label for the i-th sample (often 0 or 1 in binary classification).
- $\hat{y}_i$ is the predicted probability for the i-th sample belonging to the positive class.

The graphical representation of this Loss function can be found in Fig. 2.10.

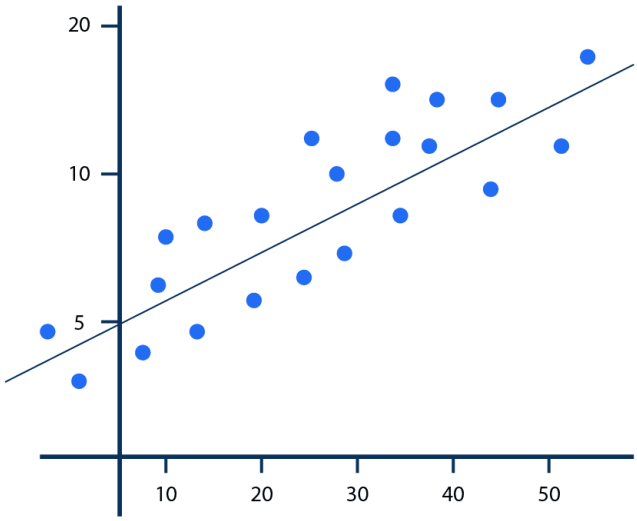


Figure 2.10. Loss function.

2.4 Convolutional Neural Networks

Computer vision is experiencing rapid development, largely driven by advances in deep learning. In this context, the concept of convolutional neural networks (CNN) emerges as an essential pillar, due to its extensive application. CNNs stand out in various computer vision applications, such as facial recognition and image classification, thanks to their ability to learn parameters such as weights and biases, similar to traditional neural networks. This similarity highlights the adaptability of CNNs to complex model for computer vision problems. The minimum necessary configuration of a CNN network contains six layers, beyond this configuration, the operation of a CNN network is impossible. This minimum configuration is based on:

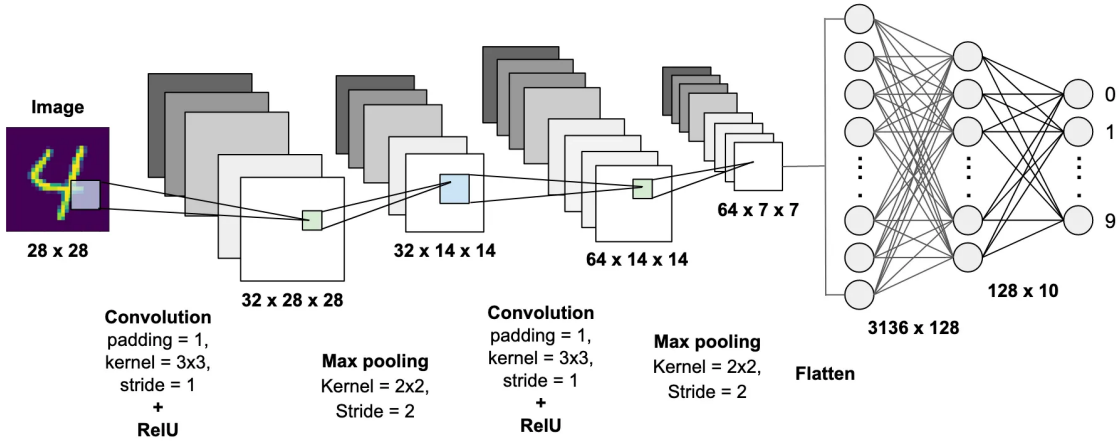


Figure 2.11. Prototype deep CNN with 2 hidden layers, 2 pooling layers, fully linked layer, and 9 output units.

Input layer of CNN: The input layer in CNN must contain data describing the signal or image to be processed. These data must be represented by their original dimensions.

Convolutional layer: In the field of image processing and deep learning, the convolution layer plays a central role, often referred to as the feature extraction layer. This name highlights its ability to identify and isolate the salient attributes of an image, thus facilitating the image analysis and recognition processes. The operation of this layer is initiated by the selection of a specific region of the image, which is then processed by a convolution operation according to the following equation:

$$x(i,j) = (\mathrm{I} * \mathrm{K})(i,j) = \sum_m \sum_n \mathrm{I}(m,n)\mathrm{K}(i-m, j-n) \tag{2.7}$$

Where:

- $x(i,j)$ is the output pixel at position (i,j) in the feature map.
- I is the input to the matrix or part of the image we are processing.

- K is the convolution kernel or filter applied to the input.
- $\sum_m \sum_n$ represents the sum over all valid positions of m and n where the kernel K is applied on the input I.
- I(m,n) is the value of the input pixel at position (m,n).
- K$(i-m, j-n)$ is the value of the convolution kernel at the adjusted position according to the applied offset from the output position (i, j).

The result is a single value that helps form the output volume, representing a map of features extracted from the image. To ensure complete coverage of the image, the filter is moved across the entire image in regular increments, called steps or strides, allowing the convolution process to be repeatedly applied to new areas. This iterative process transforms the initial image into a set of feature maps that encapsulate essential information about the visual patterns present, such as edges, textures, or shapes. Through this transformation, convolutional neural networks (CNNs) are able to perform image classification, detection, and segmentation tasks with remarkable accuracy, leveraging the extracted features to make decisions based on visual content.[18]

- **Pooling layer :** The pooling layer is used to reduce the spatial volume of the input image after convolution. It is used between two convolution layers. If we apply a fully connected (FC) layer after the convolution layer without applying pooling or max pooling, the computation will be expensive. Thus, maximum pooling is the only way to reduce the spatial volume of the input image by encoding the information. The main used functions are:
  - Max Pooling : This operation selects the maximum of values in a defined neighborhood region.

$$MaxPooling(x_{i,j}) = \max_{(p,q)\in R_{i,j}} x_{p,q} \tag{2.8}$$

  - Average Pooling : This operation calculates the average of the values in a defined neighborhood region.

$$AveragePooling(x_{i,j}) = \frac{1}{|R_{i,j}|} \sum_{(p,q)\in R_{i,j}} x_{p,q} \tag{2.9}$$

  - Sum Pooling: This operation calculates the sum of values in a defined neighborhood region.

$$SumPooling(x_{i,j}) = \sum_{(p,q)\in R_{i,j}} x_{p,q} \tag{2.10}$$

– Min Pooling: This operation selects the minimum of values in a defined neighborhood region.

$$\mathrm{M}inPooling(x_{i,j}) = \min_{(p,q)\in \mathrm{R}_{i,j}} (x_{p,q} \tag{2.11}$$

where:

$x_{i,j}$ represents the element at position (i,j) of the input matrix and $\mathrm{R}_{i,j}$ is the neighborhood region (window) around the element (i,j).[18]

- **Fully connected layer:** A fully connected layer involves weights, biases, and neurons. It connects neurons in one layer to neurons in another layer. It is used to classify images into different categories by training. This process can be defined by the mathematical equation below: The equation representing the output y of a fully connected layer is given by:

$$y = f(\mathrm{W} \cdot x + b) \tag{2.12}$$

where:

- y represents the output vector of the fully connected layer.
- W is the weight matrix of the layer.
- x is the input vector to the layer.
- b is the bias vector added to the weighted sum of inputs.
- f is the activation function applied element-wise.
- **Softmax layer :** The Softmax or logistic layer represents the final layer within a Convolutional Neural Network (CNN), situated at the terminus of the Fully Connected (FC) layer. The logistic layer is predominantly employed for binary classification tasks, while Softmax is utilized for multi classification purposes.

$$softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{2.13}$$

- **Output layer :** The output layer encapsulates the encoded label information.[18]

2.4.1 Hyperparameters

A parameter set before to the learning process. These parameters can be modified and greatly affect the model performance. Among important parameters are:

Learning rate: is a parameter regulates how much the model will alter when the model weights are updated, in response to the expected error. Choosing the

learning rate is tricky since a value that is too small may result in a protracted training cycle that may get stuck, while a value that is too high could result in learning a sub-optimal set of weights that are too fast or unstable.

Batch size: refers to the amount of samples examples utilized in the one iteration correspond to such subgroup.

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

2.4.2 Autoencoder

Autoencoders are a special type of unsupervised feedforward neural network (no labels needed). The main application of autoencoders is to accurately capture the key aspects of the provided data to provide a compressed version of the input data, generate realistic synthetic data, or flag anomalies.[28] Autoencoders are composed of 2 key fully connected feedforward neural networks (Figure 1):

- **Encoder:** compresses the input data to remove any form of noise and generates a latent space/bottleneck. Therefore, the output neural network dimensions are smaller than the input and can be adjusted as a hyperparameter in order to decide how much lossy our compression should be[28].
- **Decoder:** making use of only the compressed data representation from the latent space, tries to reconstruct with as much fidelity as possible the original input data (the architecture of this neural network is, therefore, generally a mirror image of the encoder). The "goodness" of the prediction can then be measured by calculating the reconstruction error between the input and output data using a loss function.[28]

Repeating iteratively this process of passing data through the encoder and decoder and measuring the error to tune the parameters through backpropagation, the Autoencoder can, with time, correctly work with extremely difficult forms of data.[28]

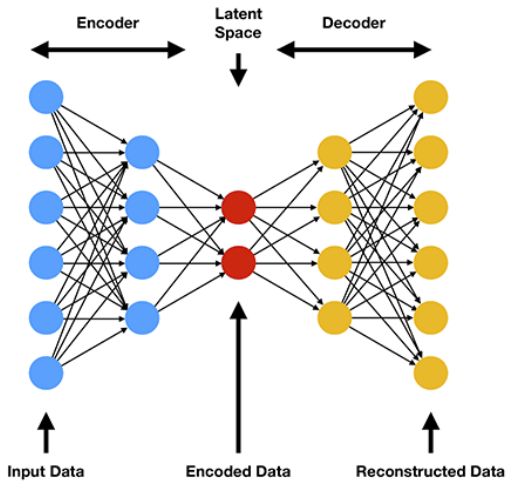


Figure 2.12. Autoencoder architecture

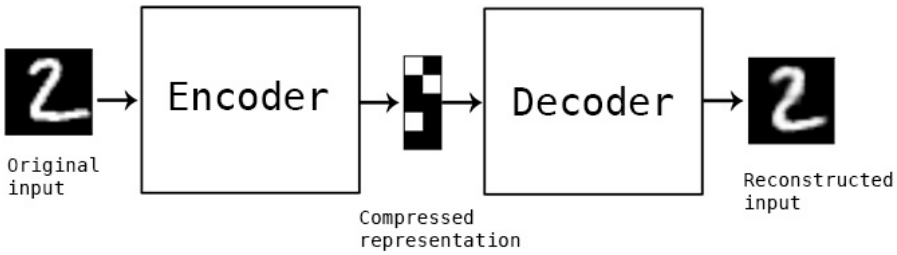


Figure 2.13. Autoencoder schema

2.4.3 Approaches to Training

There are three common methods, approaches, or strategies applied in deep learning to perform classification, and segmentation tasks:

1- Training from scratch:

Requires a very big labeled data set to be obtained, in addition, network architecture needs to be created to learn features and models. This is great for new apps or apps with a high number of output categories. This is a less popular strategy since, with a huge amount of data and a high learning rate; such networks generally take days, weeks,possible until months to train.

2- Transfer learning:

Recently, is noticed that an increasingly used to deep neural networks in many areas, thus to construct a deep learning network with an acceptant outcome, demands enormous amounts of training data, and significant training time [29]. Transfer learning is the motivated way to answer one problem from the pre-trained model. It gains the

parameters and weights from an already-trained network and utilized as the starting point on a new problem, hence change them to be fit in the target model. Transfer learning may also support the target project learning inside the following ways:

- Stepped forward baseline performance: when we raise the knowledge 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: adopting information from a source model may additionally assist in totally getting understanding of the target task, compared to learn from scratch. Therefore, improve the total time taken to construct or train a model.

3- Fine-tuning:

A procedure to train the pre-trained network on a tiny dataset. Usually, fully linked layers, which may be utilized as classification layers, the pre-trained layer will receive a lower learning rate. This will enable the characteristics to be modified to the new dataset.

2.5 Regularization Techniques:

One of the most common difficulties facing data science practitioners is to prevent overfitting. When the model did exceedingly well on train data, but was unable to predict test results. Different regularization strategies to tackle the overfitting issue such as:

- **Data Augmentation:** this is a strategy that permits expanding the data available for training models without collecting new data. This might be in different schemes dependent on the dataset used. Techniques for data augmentation, such as cropping, padding, and horizontal flipping are commonly employed to train massive neural networks.
- **Dropout:** the fundamental principle is to deactivate input units by random after each repetition. The set of 2^n neural networks is a neural network with n units that hires dropout. Such alternative networks feature a lower number of units but also swap weights in order to keep an equivalent overall number of parameters. Training a weight loss network can be understood as training a series of 2^n smaller weight share networks.
- **Early Stopping:** while training, the model is checked on dataset after each epoch. When the output of the model on the validation dataset begins to degrade such as the accuracy begins to fall or the loss begins to increase, thus, the training phase is ended.

2.6 An overview of polyp classification and segmentation techniques

This section presents an analysis of the existing research on the classification and segmentation of colorectal polyps, including the techniques used, the conducted tests, and the obtained outcomes. An image classification pipeline can be divided into two main stages: feature extraction and classification training. Most conventional feature extractors rely on pre-designed methodologies, necessitating a deep understanding of the area and significant technical expertise. Furthermore, the job becomes burdensome when numerous factors are meticulously adjusted. The earlier study on the identification of colorectal polyps primarily examined methods of feature engineering. For instance, it explored extracting features from vascular architecture [30] or utilizing scale-invariant key points selected from local patches as distinctive characteristics. In a study, they utilized two learning algorithms to classify colon polyps as a multi-stage system[31]. The purpose was to perform vessel segmentation, and then employ the results as seed for the fast performing algorithm to carry out further segmentation for the entire vessel lumen. Therefore, the features obtained from segmentation were utilized to classify the polyps. The method attained a 90% accuracy rate in correctly classifying datasets containing 56 polyps, with the ground truth histologically confirmed. One drawback of these research is that the selection of each region of interest(ROI) had to be done manually.

Table 2.1. 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.

Reference	Study Goal	Method Used	Main Findings	Limitation
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.

Reference	Study Goal	Method Used	Main Findings	Limitation
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
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

Reference	Study Goal	Method Used	Main Findings	Limitation
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

2.7 Conclusion

This chapter provided a comprehensive overview of deep learning techniques, establishing the groundwork for understanding their application in colorectal polyp detection and classification. It covered the fundamentals of machine learning, artificial neural networks (ANNs), and convolutional neural networks (CNNs), highlighting their significance in image analysis tasks. The chapter also explored key concepts such as activation functions, loss functions, and hyperparameters, which are essential for building and training effective deep learning models. By delving into these foundational concepts, this chapter has laid the groundwork for the subsequent chapters, where these techniques will be applied to address the specific challenges of colorectal polyp detection and classification.

Chapter 3

Colorectal Polyp Detection and Segmentation, DL implementation, Results and discussions

3.1 Introduction

This chapter goes into the practical implementation of the research and presents the experimental results obtained. It will cover the details about the hardware and software environment used for the experiments, including the programming language, libraries, and computational resources. It will also introduce the specific programming language (Python) and libraries (Keras, TensorFlow) employed for implementing the deep learning models and data processing pipelines. The chapter will then explain the steps taken to prepare the input data, such as patch extraction from the CVC-ClinicDB dataset and data augmentation techniques to enhance the dataset's size and diversity. Finally, it will present and analyze the results obtained from evaluating the deep learning model on the CVC-ClinicDB dataset, discussing the performance metrics of the model proposed, such as accuracy, precision, recall, and specificity, and interpreting these metrics in the context of polyp detection and classification.

3.2 Experiments Configuration

All the experiments were carried. It is that allows performing professional and advanced projects using Jupyter, which enables us to implement the entire project in Python. the hardware specifications allocated and used in this project are as follows:

- GPU NVIDIA GeForce RTX 3060 6 GB.
- Intel(R) Core(TM) i7 CPU @ 2.30GHz.
- 32GB RAM.

3.3 Programming Language and Libraries

There are various deep learning resources available that are useful for developers, which made the implementation components and procedures efficient. Several toolkits and libraries are used after careful evaluation and depending on the unique requirements and time limits of this work. The following subsections will introduce them.

3.3.1 Keras

Is a Python based open Source application programming interface that uses either Theano or Tensorflow as backend [32]. It was designed to allow rapid innovation, allowing complete solutions to be more easily created and readable with the biggest range of CNN model.

3.3.2 TensorFlow

TensorFlow is a Python open source library created and released by Google [33], 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.

3.3.3 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 multi-functional 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.

3.4 DL Implementation, Experimental Results and discussion

The proposed Deep Learning Model to achieve the detection and segmentation of colorectal polyps is presented in the flowchart bellow:

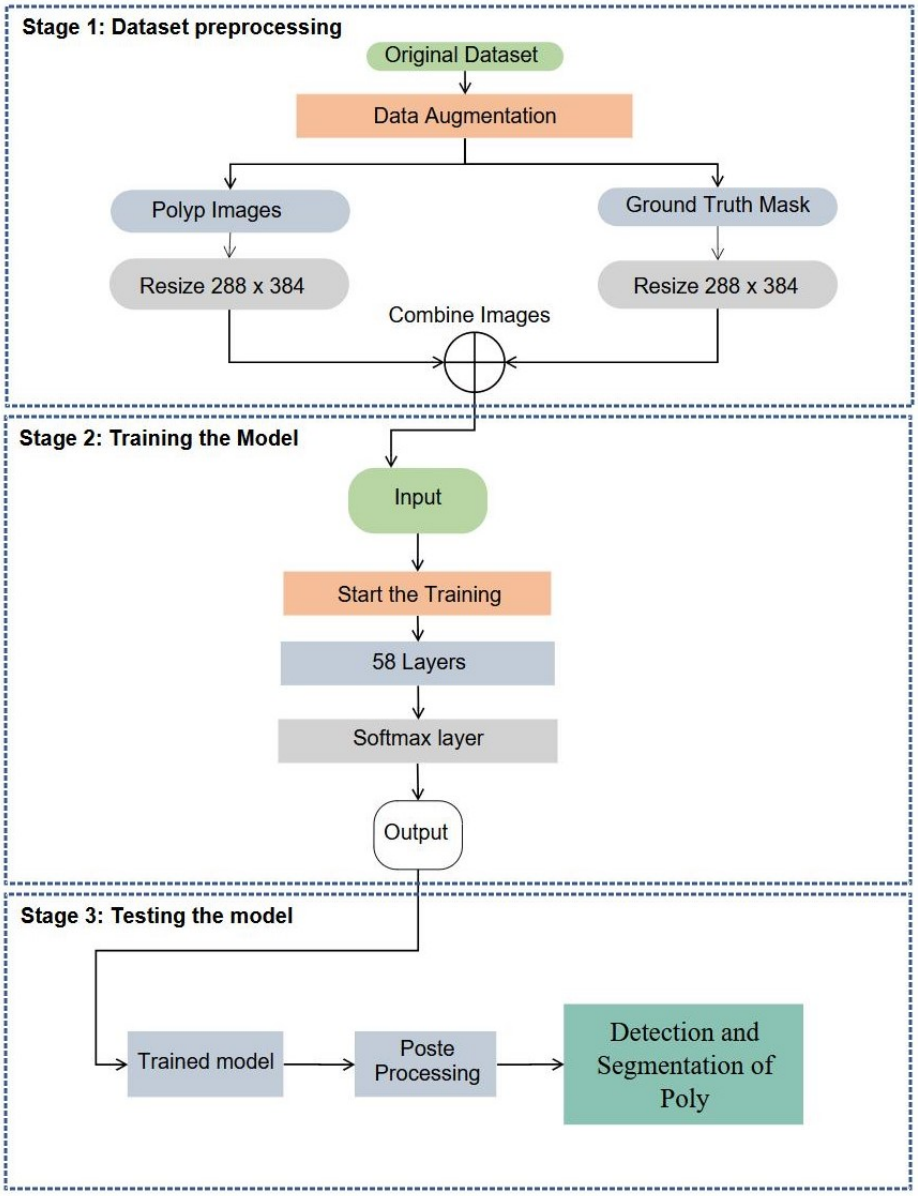


Figure 3.1. A representation of the Deep Learning model for colorectal polyp detection and classification.

3.4.1 The used Dataset

CVC-ClinicDB [34] is a database comprising frames extracted from colonoscopy videos, featuring a variety of polyps with different shapes and sizes. The dataset consists of 612 images, each accompanied by a ground truth mask outlining the polyp region.

The images were sourced from 25 distinct colonoscopy video studies, ensuring at least one polyp-containing sequence per study and resulting in a total of 29 sequences in the database. These polyp annotations, provided by medical professionals, serve as benchmarks for evaluating our proposed polyp segmentation approach.

For our experiments, a subset of 15 sequences was selected, originating from unique studies and each containing RGB color images at a resolution of 384x288 pixels. From

these sequences, frames were meticulously chosen to exhibit diverse perspectives of polyps.

Notably, this dataset was originally curated as the training set for the MICCAI2015 and ISBI2015 sub-challenges on polyp recognition in endoscopic videos. The frames were extracted from 29 endoscopic videos generously provided by Hospital Clinic, Barcelona, Spain. The dataset includes 612 standard-definition still images, each annotated with ground truth markings corresponding to the polyp region, as illustrated in Figure 3.1

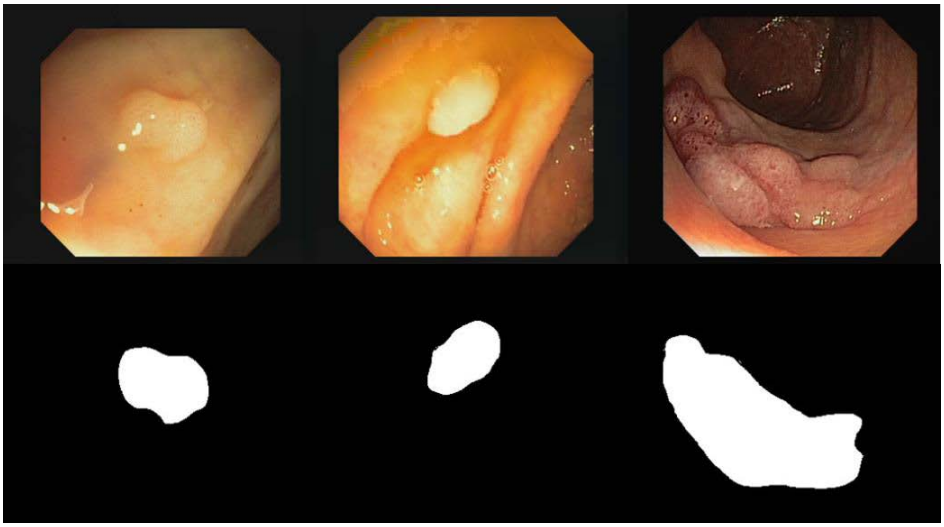


Figure 3.2. Sample images of CVC-ClinicDB. Original images shown in the first row and their corresponding ground truth shown in the second row.

3.4.2 Data Preprocessing

3.4.2.1 Data Augmentation

Data augmentation is an important technique that has been widely used in the machine learning pipeline. By introducing variations of images, such as different orientation, location, scale, brightness, etc., to existing data, we can increase the robustness and reduce over-fitting of our model. We apply some basic augmentation methods such as random rotation, random horizontal, vertical flip, and random zoom. We also apply tilt (skew), shearing, and random distortion since the tortuous appearance of the intimate is caused by the movement of the colonoscopy lens and the colon. Besides, we apply random contrast and a random brightness change to simulate different lighting environments or different photography equipment that may occur during colonoscopy procedures. Fig. 3.5 shows a sample of the augmented images after cropping, rotating, tilting, shearing, and distorting. The data augmentation is implemented by the image augmentation library 'Augmentor' [35].

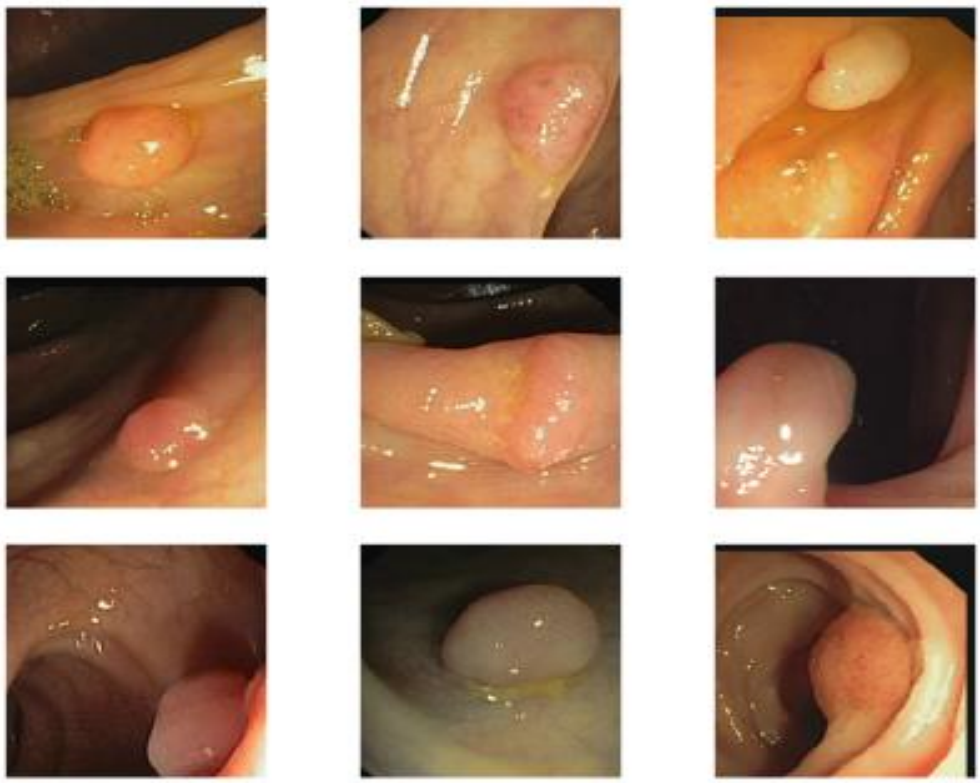


Figure 3.3. Sample augmented polyp images with different shapes and sizes.

3.4.2.2 Data splitting

The dataset have 612 images are splited into an 80 and 20 percent. 80% for training process and 20% for testing to validate the training process.

Two different major classes were utilized as targets in the training and testing stages, and they were arranged into folders that comprised two subfolders designated 0 and 1, where 0 represents a collection of Ground truth images and 1 is responsible for representing lesions defined as polyps in the image. The table 4.1 summarizes the splitting technique for training, validation, and testing samples.

Table 3.1. Training-testing split details

Method	Ratio (%)	Training	Validation
Split sample	80/20	489	122

3.4.2.3 Data resizing

we resize the input images to $384 \times 288 \times 2$, and then normalize the input images with the mean and standard deviation of the train set. The normalizing technique can make the model converge faster during the training phase since gradients can approach the lowest in a direct path.

3.4.3 The Proposed CNN deep Learning model

In this section, we introduce a description of all the necessary parts to achieve colorectal abnormalities detection and classification based on CVC-clinic database images presented in subsection 3.6.2 The proposed methodology consists of two main steps: The first focuses on the architecture to build a CNN deep learning network for colorectal abnormalities detection and classification based on 80% of the content of the used database, and the second step focuses on the validation process of the proposed architecture using the remaining 20% of the database.

The U-Net architecture is a type of convolutional neural network (CNN) specifically designed for image segmentation tasks, making it highly suitable for colorectal abnormality detection and classification projects. Its unique design allows it to capture both local and global context, leading to accurate and precise segmentations. The proposed architecture of the CNN deep learning network is U-Net based on 58 layers as below:

- **Image input layer:** 288x384x2 Image with zero center normalization.

  Encoder stages

- **2-D Convolution:** 64 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 64 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Maxpooling:** 2 x 2 max pooling with stride [2 2] and padding [0 0 0 0].
- **2-D Convolution:** 128 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 128 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Maxpooling:** 2 x 2 max pooling with stride [2 2] and padding [0 0 0 0].
- **2-D Convolution:** 256 3 x 3 convolutions with stride [1 1] and padding 'same'.

- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 256 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Maxpooling:** 2 x 2 max pooling with stride [2 2] and padding [0 0 0 0].
- **2-D Convolution:** 512 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 512 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Dropout:** 50% dropout.
- **2-D Maxpooling:** 2 x 2 max pooling with stride [2 2] and padding [0 0 0 0].

Bridge stages

- **2-D Convolution:** 1024 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 1024 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Dropout:** 50% dropout.
- **Decoder Stages**
- **2-D Convolution:** 512 2 x 2 transposed convolutions with stride [2 2] and cropping [0 0 0 0].
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Depth Concatenation:** Depth concatenation of 2 inputs.
- **2-D Convolution:** 128 6 x 6 convolutions with stride [1 1] and padding 'same'.

- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 512 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D transposed convolution:** 256 2 x 2 transposed convolutions with stride [2 2] and cropping [0 0 0 0].
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Depth concatenation:** Depth concatenation of 2 inputs.
- **2-D Convolution:** 512 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 256 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Transposed Convolution:** 128 3 x 3 transposed convolutions with stride [2 2] and cropping [0 0 0 0].
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Depth Concatenation:** Depth concatenation of 2 inputs.
- **2-D Convolution:** 128 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 128 3 x 3 convolutions with stride [1 1] and padding 'same'.
- **Relu:**Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Transposed Convolution :** 64 2 x 2 transposed convolution with stride [2 2] and cropping [0 0 0 0].

- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **Depth Concatenation:** Depth concatenation of 2 inputs.
- **2-D Convolution:** 64 3 x 3 convolution with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** 64 3 x 3 convolution with stride [1 1] and padding 'same'.
- **Relu:** Sets all negative values in the input to zero and leaves positive values unchanged.
- **2-D Convolution:** it's the final-Convolution Layer 2 1 x 1 convolution with stride [1 1] and padding 'same'.
- **Softmax:** Converts the raw output scores of the network into probabilities that sum up to 1, making it easier to interpret the final output as class probabilities.
- **Pixel classification Layer (Output):** segmentation-Layer,it based on Cross-entrpy loss.

As technical view, the U-net architecture is realy complicated model with big number of layers, we can illustrate this architecture by the synoptic scheme as shown in the figure bellow:

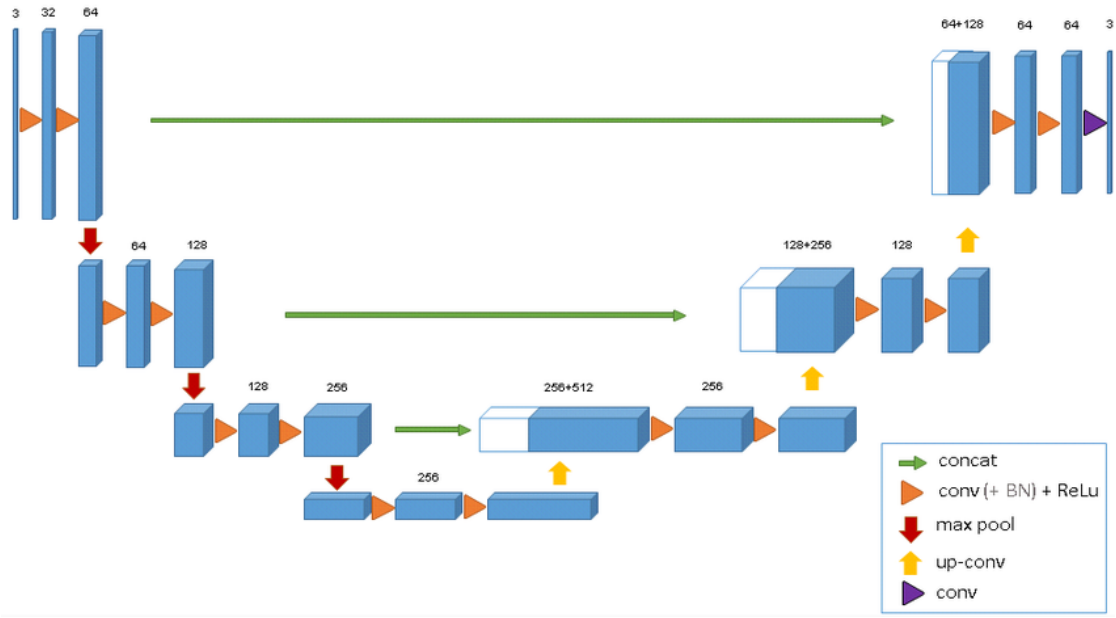


Figure 3.4. U-net synoptic model.

3.4.4 Training and testing process

The proposed model is trained and tested using a single NVIDIA GeForce RTX 3060 GPU. We use the binary-cross-entropy loss function to calculate the loss for each pixel over the final output segmentation map.

3.5 Obtained results and discussion

3.5.1 Evaluation metrics

3.5.1.1 Dice Coefficient and Jaccard:

is a spatial overlap index and a repeatability validation parameter used in machine learning, especially in semantic image segmentation. It evaluates the similarity between the prediction binary segmentation result and the ground truth mask. The value of a dice coefficient ranges from 0, indicating no spatial overlap between two sets of binary segmentation results, to 1, indicating total overlap. The value of the Dice coefficient equals twice the number of items common to both sets divided by the sum of the number of elements in each set. where |S| and |Tr| are the cardinalities of the two sets (the number of pixels in each binary mask image). To analyze and evaluate the performance of the CNN model from the perspective of machine learning, all the results of the Dice coefficient are calculated before the post processing. Evaluating the performance of segmentation criteria is crucial for determining the effectiveness and accuracy of proposed segmentation methods in image analysis. The commonly used criteria by researchers to evaluate the quality of image segmentation methods include:[18][36]

$$\mathrm{D}ice = \frac{2|\mathrm{S} \cap \mathrm{T}r|}{|\mathrm{S}| + |\mathrm{T}r|} \tag{3.1}$$

$$Jaccard = \frac{\mathrm{S} \cap \mathrm{T}r}{\mathrm{S} \cup \mathrm{T}r} \tag{3.2}$$

Where:

- S: Ground truth segmentation region.
- Tr: Predicted segmentation region.

True Positive (TP): The number of positive samples that are identified correctly by the classifier represents that the sample falls into the polyp class and is classed as such.

False Positive (FP): The number of negative samples that are incorrectly recognized in a positive category signifies that sample falls in the non-polyp class but is labeled as polyp class.

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

False Negative (FN): The number of positive samples that are erroneously recognized in another category signifies that the sample falls in the polyp class but is classed as non-polyp.

Confusion matrix: is a table utilized to describe the overall performance of the classification model on test data whose true values are known. The correlations between true positive, false positive, true negative, and false negative are presented in Table 3.1.

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

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

3.5.1.2 Recall

Calculates the proportion of all true positive samples from cases that are actually positive. Also it is referred to as sensitivity and true positive rate.

$$\text{Recall} = \frac{\text{True Positive}}{\text{Total Actual Positive}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.4}$$

3.5.1.3 Precision

Calculates the proportion of all true positive samples from cases that are predicated as positive.

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

3.5.1.4 Accuracy

Calculates the proportion of correctly classified samples.

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

3.5.1.5 F1 score

Another accuracy measure, also referred to as F-measure, utilized to seek the relation between precision and recall by counting the weighted average.

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

3.5.1.6 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 the true positive rate (TPR) represents the y-axis at various threshold settings.

3.5.1.7 Specificity

Calculates the proportion of all true negative samples from cases that are actually negative, also referred to as false positive rate.

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

3.5.2 Obtained results and discussion

3.5.2.1 Global obtained results

The results of the U-Net model evaluation using the CVC-ClinicDB dataset are impressive. The figures bellow present an examples of polyp coloretal detection and segmentation using the proposed model:

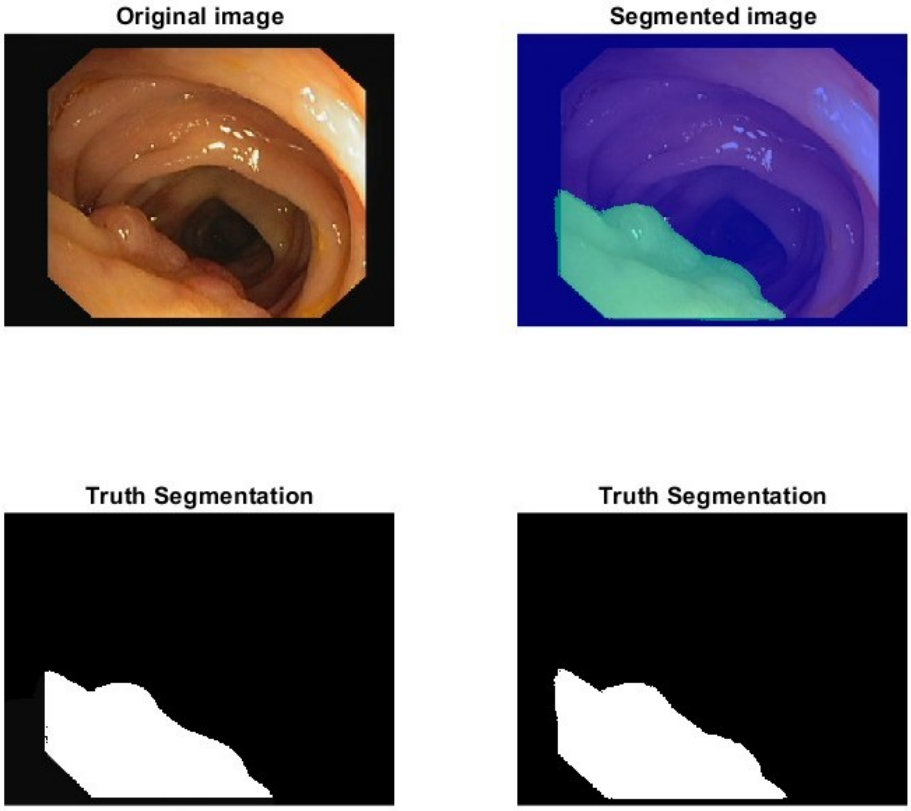


Figure 3.5. Polyp colorectal detection and of segmented (images index 1)

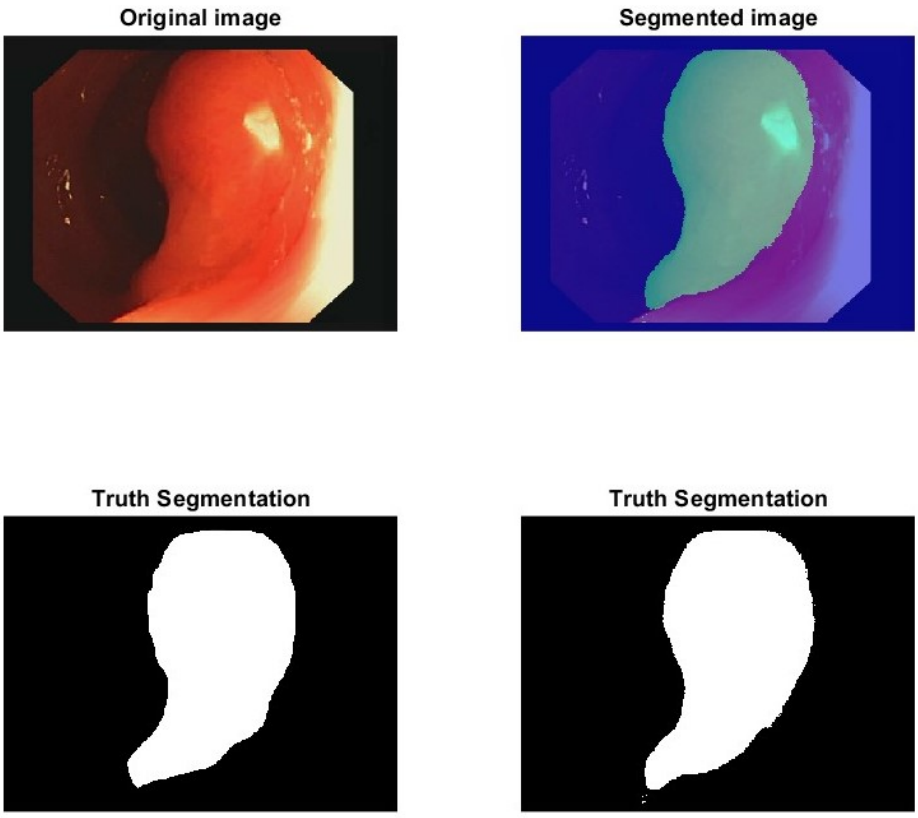


Figure 3.6. Polyp colorectal detection and of segmented (images index 320)

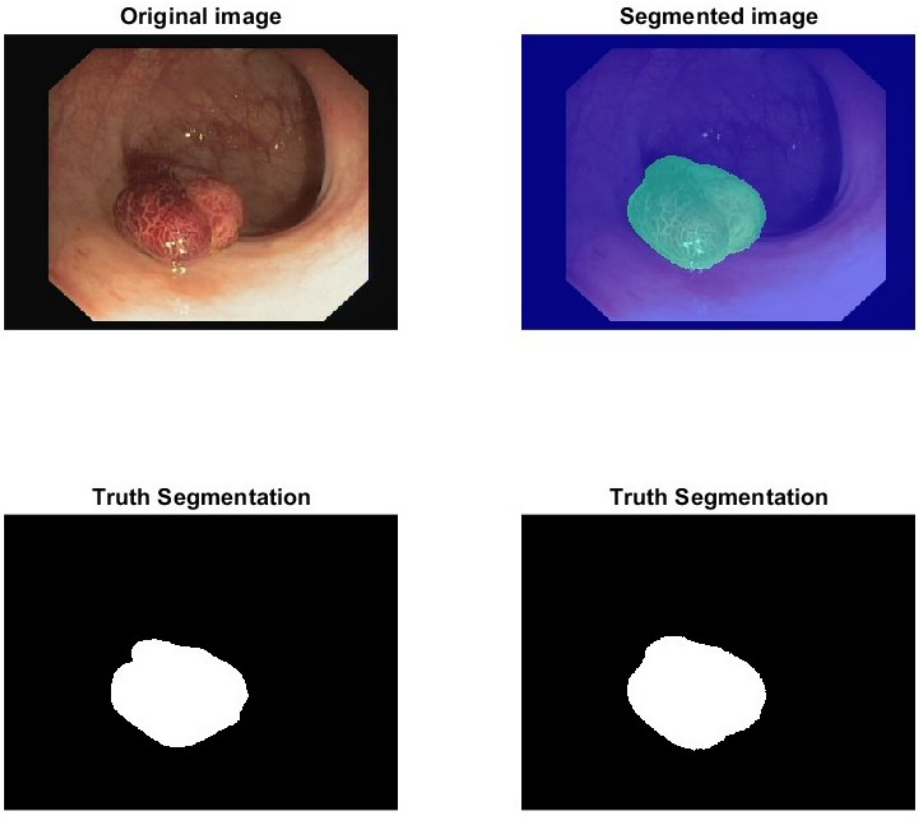


Figure 3.7. Polyp colorectal detection and of segmented (images index 470)

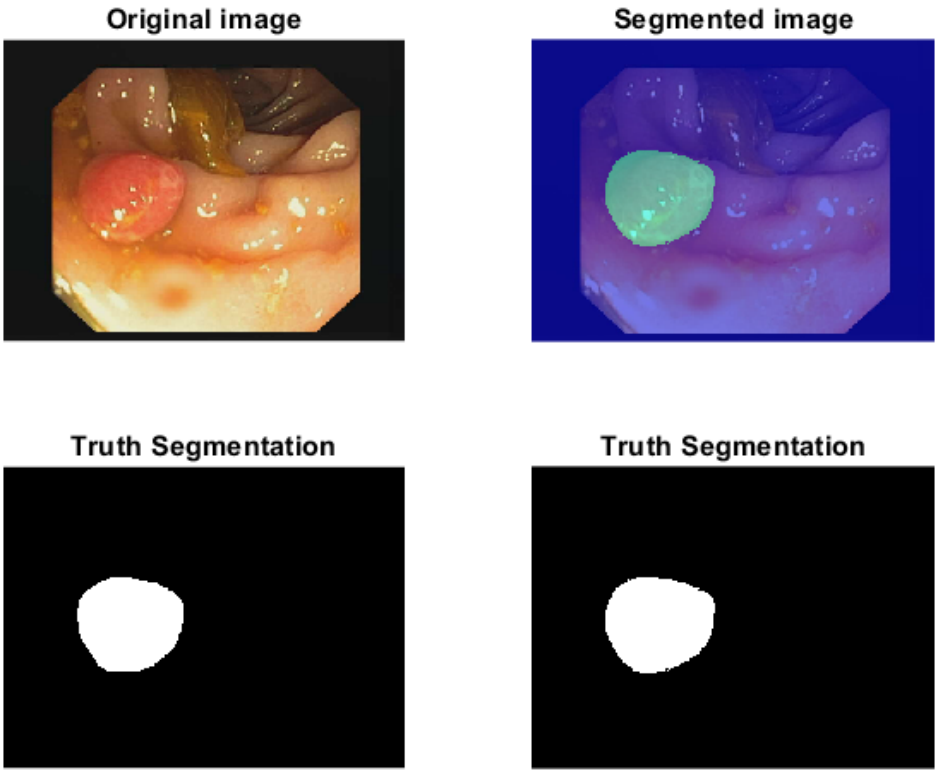


Figure 3.8. Polyp colorectal detection and of segmented (images index 612)

By analyzing all the images of CVC-ClinicDB, the obtained metrics results are presented in the table bellow:

Table 3.3. Metric Values (Mean ± Std Dev)

Metric	Value (Mean ± Std Dev)
Accuracy	0.9867 ± 0.0142
Precision	0.9867 ± 0.0142
Sensitivity (Recall)	0.9867 ± 0.0142
Specificity	0.9866 ± 0.0142
Dice Coefficient	0.5000 ± 0.0001

The obtained results showing the metrics (Accuracy, Precision, Sensitivity/Recall, and Specificity) for 612 evaluated images.

The model achieved a mean accuracy of 98.67% across all 612 images. It means correctly classified approximately 98.67% of the pixels in the images as either belonging to the segmented region or the background. The standard deviation of 1.42% indicates some variability in performance across images. The mean precision of 98.67% suggests that when the model predicts a pixel as part of the segmented region, it's correct about 98.67% of the time. This is a good indicator of the model's reliability in identifying the region of interest.

With a mean sensitivity of 98.67%, the model successfully detected approximately 98.67% of the actual pixels belonging to the segmented region. This demonstrates the

model's ability to capture most of the relevant structures in the images.

The mean specificity of 98.66% indicates that the model correctly identified about 98.66% of the background pixels. This is important for minimizing false positives, where background regions are incorrectly labeled as part of the segmented region.

As Dice coefficient metric value, it around 0.5 for all images. In the context of image segmentation, a Dice coefficient of 0.5 indicates a moderate overlap between the predicted segmentation and the ground truth. This suggests that while the model is generally accurate, there's room for improvement in terms of precisely matching the boundaries of the segmented regions.

In addition, the very low standard deviation (0.0001) indicates that the model's performance is highly consistent across all images in terms of the Dice coefficient. This consistency is a positive aspect, as it suggests that the model's predictions are reliable and not prone to large fluctuations.

3.5.2.2 Best and Worst obtained results:

To achieve an accurate evaluation of the proposed model, we suggest to highlight the best and the worst obtained results, these later are presented in the tables bellow:

Table 3.4. Summary of Dice Coefficient Metrics

Metric	**Value**
Mean Dice Coefficient	0.5000
Standard Deviation	0.0001
Best Dice Coefficient	0.5025 (Image Index 154)
Worst Dice Coefficient	0.4986 (Image Index 155)

Table 3.5. Summary of Jaccard Index Metrics

Metric	**Value**
Mean Jaccard Index	0.3333
Standard Deviation	0.0001
Best Jaccard Index	0.3356 (Image Index 154)
Worst Jaccard Index	0.3321 (Image Index 155)

According to the table 3.4, we notice that the best obtaining results for the image index 306, that achieved an accuracy of 99.66% (Figure 3.9), showcasing the model's potential for highly accurate segmentation.

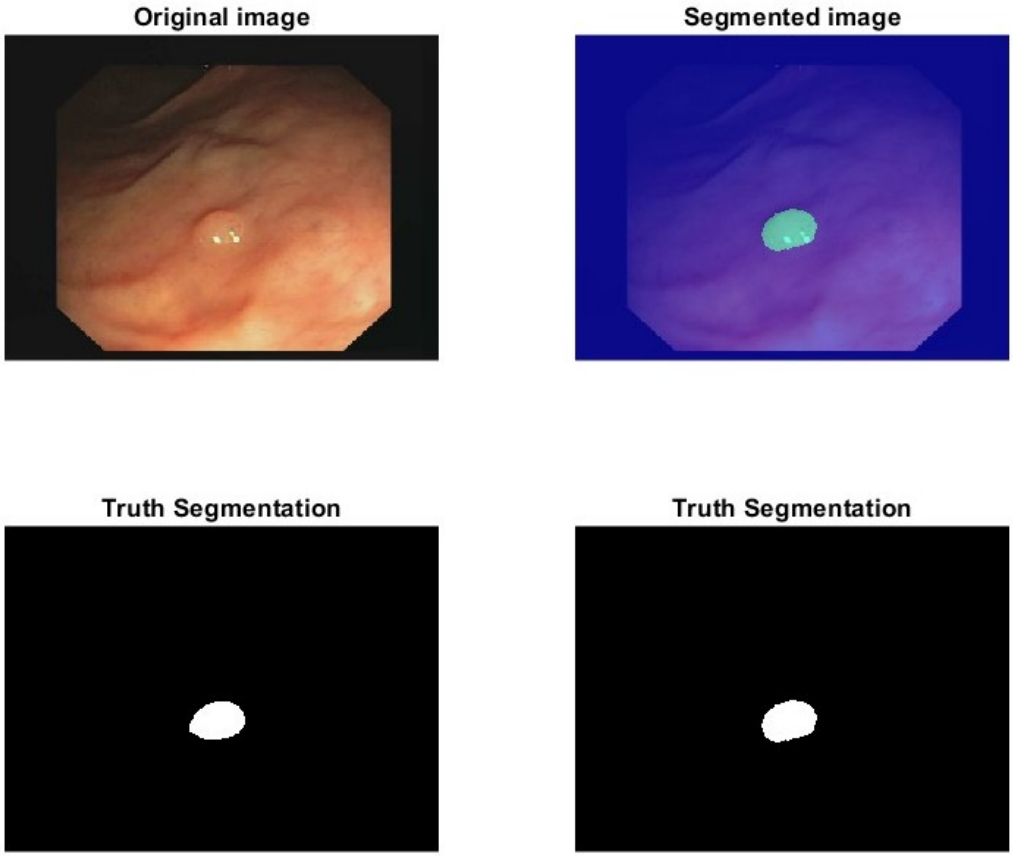


Figure 3.9. Best obtained results for image 306

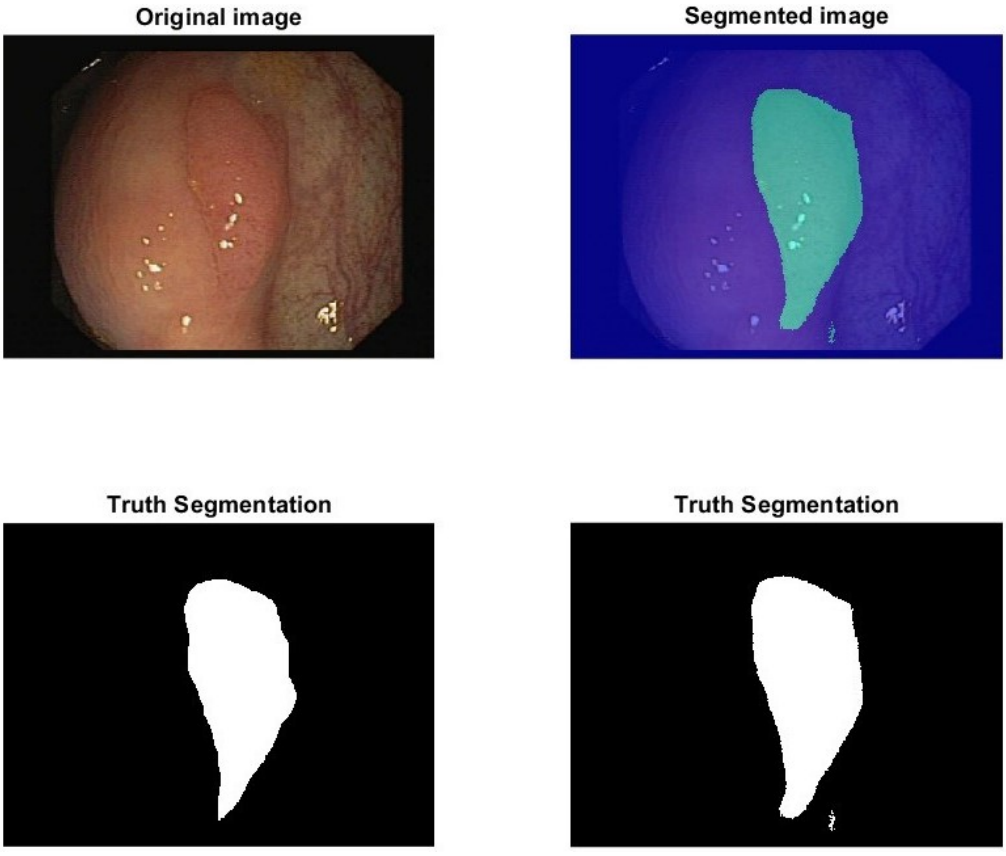


Figure 3.10. Worst obtained results for image 155

In contrast, the worst performing image (index 155) had an accuracy of 95.54% (figure 3.10). Examining the image with the lowest accuracy (95.54%) and comparing it to its corresponding ground truth segmentation can provide valuable insights into the types of challenges the model may encounter.

Potential factors contributing to decreased performance in this specific instance could:

- include poor image quality (blurriness or low contrast),
- a polyp with an unusual shape or appearance,
- a polyp located in a difficult-to-segment area (e.g., near folds in the colon),
- or even errors
- or inconsistencies in the ground truth annotation.

Further investigation of such anomalous cases revealed similar potential contributing factors, including variations in polyp morphology, challenging anatomical locations, and inconsistencies in ground truth annotations. These findings underscore the importance of addressing these abnormalities in future iterations of model training and dataset curation to enhance the overall segmentation performance and robustness of the model.

According to the table 3.5, we notice that the Jaccard index values are consistently around 0.3333 for all images. In the context of image segmentation, a Jaccard index of 0.3333 indicates a fair amount of overlap between the predicted segmentation and the ground truth. This suggests that while the model is generally accurate, there's room for improvement in terms of precisely matching the boundaries of the segmented regions.

The very low standard deviation (0.0001) indicates that the model's performance is highly consistent across all images in terms of the Jaccard index. This consistency is a positive aspect, as it suggests that the model's predictions are reliable and not prone to large fluctuations.

3.6 Conclusion

The evaluation and testing of the U-Net model on the CVC-ClinicDB dataset proved exceptional results, focusing on its capabilities in segmenting image tasks. The model achieved high average accuracy, precision, sensitivity, and specificity, confirming its ability to accurately classify pixels and detect regions of interest. The Dice coefficient and Jaccard index values across images further confirm the model's reliability and robustness. While there is a need for improvement in boundary precision, the U-Net model's performance on the CVC-ClinicDB dataset is excellent, indicating its potential for real-world applications in polyp detection and segmentation.

General Conclusion

In this dissertation, we have explored the application of deep learning, particularly Convolutional Neural Networks (CNNs), for the detection and classification of colorectal polyps in medical images. Given the increasing prevalence and mortality rates of colorectal cancer worldwide, the development of accurate and efficient methods for polyp detection is of paramount importance.

Our research focused on leveraging the power of CNNs to analyze colonoscopy images and identify regions containing polyps. We investigated various CNN architectures, models and training strategies to find the optimized model's performance, especially in the biomedical field. The results obtained from our experiments on the CVC-ClinicDB dataset were promising, demonstrating the effectiveness of our approach in accurately detecting and classifying polyps.

The high accuracy, precision, sensitivity, and specificity achieved by the model proposed highlight its potential for real-world clinical applications. By automating the polyp detection process, we can reduce the reliance on manual inspection, which is often time-consuming and prone to human error. This can lead to earlier detection of polyps, enabling timely intervention and potentially preventing the progression to colorectal cancer.

While our research has made significant strides in the field of polyp detection, there are still areas for further exploration and improvement. Future work could focus on refining the segmentation algorithms to achieve even higher precision in delineating polyp boundaries. Additionally, incorporating larger and more diverse datasets could enhance the model's robustness and generalizability.

In conclusion, this dissertation has demonstrated the potential of deep learning-based approaches for the detection and classification of colorectal polyps. The proposed model holds promise for improving the accuracy and efficiency of colonoscopy procedures, ultimately contributing to the early detection and prevention of colorectal cancer.

Bibliography

[1] S. Tchoketch Kebir, “Développement d’algorithmes de super-résolution en utilisant la transformée en ondelettes,” Ph.D. dissertation, Ecole Nationale Polytechnique, 2011. 4, 5, 6

[2] S. Tchoketch Kebir, “Développement de techniques de segmentation des images IRM pour la détection des tumeurs cérébrales,” Ph.D. dissertation, USTHB, 2019. 4

[3] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, “Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods,” *International journal of cancer*, vol. 144, no. 8, pp. 1941–1953, 2019. 7

[4] I. CiteDrive, “What is colorectal cancer,” 2024, https://www.cancer.org/cancer/ colon-rectal-cancer/about/what-is-colorectal-cancer.html [Accessed: (01-06-2024)]. 7, 10

[5] R. J. Groff, R. Nash, and D. J. Ahnen, “Significance of serrated polyps of the colon,” *Current gastroenterology reports*, vol. 10, no. 5, pp. 490–498, 2008. 7 [6] E. Dekker and D. K. Rex, “Advances in crc prevention: screening and surveillance,” *Gastroenterology*, vol. 154, no. 7, pp. 1970–1984, 2018. 7

[7] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global patterns and trends in colorectal cancer incidence and mortality,” *Gut*, vol. 66, no. 4, pp. 683–691, 2017. 7

[8] D. A. Lieberman, D. K. Rex, S. J. Winawer, F. M. Giardiello, D. A. Johnson, and T. R. Levin, “Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the us multi-society task force on colorectal cancer,” *Gastroenterology*, vol. 143, no. 3, pp. 844–857, 2012. 8

[9] D. K. Rex, C. S. Cutler, G. T. Lemmel, E. Y. Rahmani, D. W. Clark, D. J. Helper, G. A. Lehman, and D. G. Mark, "Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies," *Gastroenterology*, vol. 112, no. 1, pp. 24–28, 1997. 8

[10] C. J. Kahi, "How does the serrated polyp pathway alter crc screening and surveillance?" *Digestive diseases and sciences*, vol. 60, pp. 773–780, 2015. 8

[11] D. Taha, "Deep learning model for colorectal polyps classification deepcpc," Ph.D. dissertation, Middle East University, 2020. 9, 10

[12] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2013. 11

[13] A. Horváth, S. Spindler, M. Szalay, and I. Rácz, "Preprocessing endoscopic images of colorectal polyps," *Acta Technica Jaurinensis*, vol. 9, no. 1, pp. 65–82, 2016. 11

[14] A. N. Srivastava and J. Han, *Machine learning and knowledge discovery for engineering systems health management*. CRC press, 2016. 11

[15] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014. 14

[16] M. Awad and R. Khanna, *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer nature, 2015. 14

[17] A. Burkov, *The hundred-page machine learning book*. Andriy Burkov Quebec City, QC, Canada, 2019, vol. 1. 14

[18] S. T. Kebir, F. Berrhail, and F. Didi, "An efficient tomato diseases detection and classification methodology using cnn deep learning network," *Brazilian Journal of Technology*, vol. 7, no. 2, pp. e69 316–e69 316, 2024. 15, 24, 25, 43

[19] T. Michaeli, Y. C. Eldar, and G. Sapiro, "Semi-supervised single-and multi-domain regression with multi-domain training," *Information and Inference: A Journal of the IMA*, vol. 1, no. 1, pp. 68–97, 2012. 15

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. 15

[21] A. C. Neves, I. González, J. Leander, and R. Karoumi, "A new approach to damage detection in bridges using machine learning," in *Experimental Vibration Analysis for Civil Structures: Testing, Sensing, Monitoring, and Control 7*. Springer, 2018, pp. 73–84. 16

[22] C. Perera and M. Fernando, "Comparison of multiple linear regression and artificial neural network models for the prediction of solid waste generation in sri lanka." *Trans. Mach. Learn. Data Min.*, vol. 13, no. 1, pp. 3–25, 2020. 17

[23] I. CiteDrive, "The essential guide to neural network architectures. v7." 2024, April 10, https://v7labs.com/blog/neural-network-architectures-guide. 18, 19

[24] Dremio, "Relu activation function," June 19, 2024, https://www.dremio.com/wiki/relu-activation-function. 20

[25] "What is the tanh activation function?" 2024, https://www.educative.io/answers/what-is-the-tanh-activation-function. 20

[26] H. Phillips, "A simple introduction to softmax," 2024, https://medium.com/@hunter-j-phillips/a-simple-introduction-to-softmax-287712d69bac. 21

[27] C. Pere, "What are loss functions?" 2024, https://towardsdatascience.com/what-is-loss-function-1e2605aeb904. 22

[28] P. P. Ippolito, "Introduction to autoencoders: From the basics to advanced applications in pytorch," 2024, https://www.datacamp.com/tutorial/introduction-to-autoencoders. 26

[29] D. Sarkar, R. Bali, and T. Ghosh, *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing Ltd, 2018. 27

[30] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-aided colorectal tumor classification in nbi endoscopy using local features," *Medical image analysis*, vol. 17, no. 1, pp. 78–100, 2013. 29

[31] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf, "Classification of colon polyps in nbi endoscopy using vascularization features," in *Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260. SPIE, 2009, pp. 774–785. 29

[32] K. Team, "Keras documentation: About keras," *Keras. io*, vol. 297, 2020. 34

[33] I. CiteDrive, "Why tensorflow," 2024, https://www.tensorflow.org/about [Accessed: (01-06-2024)]. 35

[34] J. Bernal, N. Tajkbaksh, F. J. Sanchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham *et al.*, "Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015

endoscopic vision challenge," *IEEE transactions on medical imaging*, vol. 36, no. 6, pp. 1231–1249, 2017. 36

[35] M. D. Bloice, P. M. Roth, and A. Holzinger, "Biomedical image augmentation using augmentor," *Bioinformatics*, vol. 35, no. 21, pp. 4522–4524, 2019. 37

[36] S. T. Kebir and S. Mekaoui, "An efficient methodology of brain abnormalities detection using cnn deep learning network," in *2018 International Conference on Applied Smart Systems (ICASS)*. IEEE, 2018, pp. 1–5. 43

الملخص

تهدف هذه المذكرة إلى عرض تقنيات التعلم العميق، وتحديداً الشبكات العصبية التلافيفية (CNNs)، للكشف عن الأورام الحميدة في القولون والمستقيم وتصنيفها في الصور الطبية. ويُعد سرطان القولون والمستقيم مصدر قلق صحي عالمي كبير، ويُعد الكشف المبكر عن السلائل أمرًا بالغ الأهمية للعلاج الفعال والوقاية. يركز البحث على تطوير نموذج قائم على شبكة CNN لتحليل صور تنظير القولون وتحديد المناطق التي تحتوي على الزوائد اللحمية بدقة. يتم التحقيق في بنيات واستراتيجيات تدريب مختلفة لشبكة CNN، بما في ذلك التعلم التحوّلي والضبط الدقيق، لتحسين أداء النموذج. يتم تقييم النموذج على مجموعة بيانات CVC-ClinicDB،وأظهرت النتائج درجة عالية من الدقة والدقة والحساسية والخصوصية في اكتشاف السلائل وتصنيفها. يساهم هذا البحث في تطوير أدوات التشخيص بمساعدة الحاسوب لفحص سرطان القولون والمستقيم، ولديه القدرة على تحسين نتائج المرضى من خلال الكشف والتدخل المبكر.

الكلمات المفتاحية: التعلم العميق، الشبكات العصبية التلافيفية، الأورام الحميدة في القولون والمستقيم، تحليل الصور الطبية، تنظير القولون، التشخيص بمساعدة الكمبيوتر.

Abstract

This dissertation aims to present deep learning techniques, specifically Convolutional Neural Networks (CNNs), for the detection and classification of colorectal polyps in medical images. Colorectal cancer is a significant global health concern, and early detection of polyps is crucial for effective treatment and prevention. The research focuses on developing a CNN-based model to analyze colonoscopy images and accurately identify regions containing polyps. Various CNN architectures and training strategies are investigated, including transfer learning and fine-tuning, to optimize the model's performance. The model is evaluated on the CVC-ClinicDB dataset, and the results demonstrate high accuracy, precision, sensitivity, and specificity in polyp detection and classification. The research contributes to the advancement of computer-aided diagnosis tools for colorectal cancer screening and has the potential to improve patient outcomes through earlier detection and intervention.

Keywords: Deep Learning, Convolutional Neural Networks, Colorectal Polyps, Medical Image Analysis, Colonoscopy, Computer-Aided Diagnosis.

Résumé

L'objectif de ce mémoire est d'exposer les techniques d'apprentissage profond. , en particulier les réseaux de neurones convolutifs (CNN), pour la détection et la classification des polypes colorectaux dans les images médicales. Le cancer colorectal est un problème de santé mondial important, et la détection précoce des polypes est cruciale pour un traitement et une prévention efficace. La recherche se concentre sur le développement d'un modèle basé sur CNN pour analyser les images de coloscopie et identifier avec précision les régions contenant des polypes. Différentes architectures de CNN et stratégies d'entraînement sont étudiées, y compris le transfert d'apprentissage et l'ajustement fin, pour optimiser les performances du modèle. Le modèle est évalué sur l'ensemble de données CVC-ClinicDB, et les résultats démontrent une grande précision, précision, sensibilité et spécificité dans la détection et la classification des polypes. La recherche contribue à l'avancement des outils de diagnostic assisté par ordinateur pour le dépistage du cancer colorectal et a le potentiel d'améliorer les résultats des patients grâce à une détection et une intervention précoces.

Mots-clés: Apprentissage Profond, Réseaux de Neurones Convolutifs, Polypes Colorectaux, Analyse d'Images Médicales, Coloscopie, Diagnostic Assisté par Ordinateur.