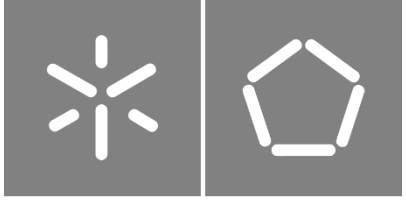
Universidade do Minho
Escola de Engenharia

Pedro Miguel Pereira Vieira

Hierarchical Classification of Lesions in Wireless Capsule Endoscopy Exams

março de 2021

Hierarchical Classification of Lesions in Wireless Capsule Ensocopy Exams
Pedro Miguel Pereira Vieira
UMinho | 2021



Universidade do Minho
Escola de Engenharia

Pedro Miguel Pereira Vieira

Hierarchical Classification of Lesions in Wireless Capsule Endoscopy Exams

Tese de Doutoramento
Programa Doutoral em Engenharia Biomédica

Trabalho efetuado sob orientação do
Professor Doutor Carlos Manuel Gregório Santos Lima
e da
Professora Doutora Carla Rolanda Rocha Gonçalves

março de 2021

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

Atribuição - Não Comercial

CC BY-NC

https://creativecommons.org/licenses/by-nc/4.0/

Acknowledgments

Antes de mais gostaria de agradecer ao meu orientador, o Professor Carlos Lima. As suas ideias e o seu apoio para levar este trabalho avante foram essenciais para os resultados obtidos.

Gostaria ainda de agradecer à equipa do Hospital de Braga por todo o apoio que deram. À Dra. Carla Rolanda, Dr. Bruno Gonçalves, Dra. Dalila Costa e Dra. Catarina Silva, obrigado pelos esclarecimentos médicos, pelo fornecimento das imagens e vídeos e por me terem ajudado a perceber que o trabalho conjunto de engenheiros e médicos afinal pode ser muito fácil.

Aos camaradas do laboratório, um grande obrigado. É impossível descrever o quão fácil foi trabalhar todos os dias sabendo que vos tinha a vocês para me aturar. Uma palavra especial ao Nuno, com quem tive a sorte de trabalhar de muito perto durante esta jornada. Tenho a certeza de que o teu futuro será cheio de sucessos.

A todos os meus amigos, principalmente ao fantástico grupo da Bola de Queijo. É impossível nomear todos, mas seria injusto se não agradecesse especialmente ao Sérgio e ao Zezé, que me acompanharam de perto nesta jornada desde o primeiro momento, e que felizmente posso considerá-los como grandes amigos. Tenho a certeza que não teria chegado a bom porto se não fosse por vocês.

À minha família. É impensável não agradecer a toda a minha pequena grande família: avós, tios e primos. Mas principalmente agradeço aos meus pais. Um obrigado pelos ensinamentos e por me terem apoiado em todas as decisões que tomei. Ao meu irmão João, que foi e é um orgulho, e que o seu exemplo me ajudou a traçar este caminho. Mas principalmente à minha irmã Paula, obrigado pelas nossas conversas intermináveis, pela presença constante e pelo apoio incondicional.

Por último, e porque sei que ele está a ler este texto, o meu maior obrigado ao Zé. Foste e és um pilar na minha vida, e tenho a mais absoluta certeza de que a tua presença nos últimos anos foi essencial para que esta etapa chegasse ao fim. É impossível descrever por palavras a importância que tens na minha vida e na pessoa que sou hoje. Que a nossa vida continue a brilhar sempre.

The author was funded by the grant SFRH/BD/92143/2013 from the Portuguese Foundation for Science and Technology (FCT), with funds from the European Social Fund (FSE), under the Human Capital Operational Programme (POCH) from Portugal 2020 Programme.

Statement of Integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, _____

Pedro Miguel Pereira Vieira

Signature: _____

Classificação Hierárquica de Lesões em Exames de Cápsula Endoscópica

Resumo

A cápsula endoscópica é um dispositivo médico que tem como como principal vantagem a possibilidade de visualizar todo o trato gastrointestinal. Este exame não invasivo é especialmente usado e vantajoso para o diagnóstico de patologias do intestino delgado, já que a endoscopia convencional é um exame invasivo que não possibilita a visualização deste órgão. Para analisar os exames de cápsula endoscópica o pessoal médico necessita de treino especializado, tendo sido provado que a quantidade massiva de imagens de cada exame pode levar à existência de erros médicos e uma propensão a que exista um subdiagnóstico de algumas patologias.

Esta tese teve como objetivo o desenvolvimento de sistemas de deteção automática de diferentes tipos de lesões presentes no intestino delgado. Estes métodos envolveram o uso de algoritmos de segmentação baseados em métodos probabilísticos (nomeadamente o *Expectation-Maximization*), com a apresentação de um método de aceleração da convergência do algoritmo e do desenvolvimento de um novo método para melhorar as fronteiras de segmentação, baseado em Campos Aleatórios de Markov. Além disso, foram estudadas diferentes metodologias de classificação supervisionada, desde classificadores mais simples e classificadores ensemble para deteção de lesões individuais, e redes neuronais convolucionais e segmentação de instâncias para deteção e segmentação de multi-patologias. Com o apoio do Hospital de Braga, foi efetuado um estudo clínico com o método desenvolvido para deteção automática de angioectasias. Este trabalho teve como principal objetivo comparar a eficiência e performance deste método com a performance de diferentes médicos a analisar exames de cápsula endoscópica.

Os diferentes métodos desenvolvidos demonstraram resultados superiores aos encontrados na bibliografia mais recente. É importante referir que o trabalho desenvolvido nesta tese permitiu uma melhor análise à necessidade de uma maior implantação de métodos de deteção de lesões em sistemas de cápsula endoscópica, tal como a necessidade de maiores e melhores estudos clínicos, tal como a disponibilização de melhores bases de dados públicas.

Palavras-Chave: Cápsula Endoscópica; Deteção Automática de Lesões; Segmentação; Segmentação de Instâncias.

Hierarchical Classification of Lesions in Wireless Capsule Endoscopy Exams

Abstract

The wireless capsule endoscopy is a medical device with the main advantage of being able to visualize the whole gastrointestinal tract. This non-invasive exam is specially used for the diagnosis of small bowel pathologies, since the conventional endoscopy is not able to visualize this organ. To analyze these exams the medical staff need specialized training and it was recently proven that the massive quantity of images that are generated lead to medical errors and consequently the sub diagnosis of certain pathologies.

In this thesis the main objective was to develop systems for automatic detection of different lesions present in the small bowel. These developments included the use of segmentation algorithms based on probabilistic methods (namely the Expectation-Maximization), with the presentation of an acceleration method and a new approach for improving the borders of the segmentation based on Markov Random Fields. Beyond that, several supervised classification strategies were studied, with the use of single-based classifiers and ensemble-based classifiers for detection of single lesions and convolutional neural networks, and instance segmentation for multi-pathology detection and segmentation. With the support of Hospital of Braga, a clinical studied was performed with the developed method for angioectasia detection. This work had the main purpose of comparing the efficiency and performance of the method with the performance of different physicians when analyzing wireless capsule endoscopy exams.

The developed methods were tested in different applications and it was found that the performance was improved when compared to the most recent bibliography. It is important to state that all this work allowed to conclude that these systems need to have a greater implantation in the clinical practice. While there is a lot of advances in computer vision methods for lesion detection, there are still lacking better clinical studies and better and bigger public databases to improve the testing of the methodologies.

Keywords: Capsule Endoscopy; Automatic Lesion Detection; Segmentation; Instance Segmentation.

Table of Contents

Index of Figures

Index of Tables

List of Abbreviations

AUC	Area Under the ROC Curve
AP	Average Precision
CAD	Computer Aided Diagnostic
CT	Computed Tomography
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DCT	Discrete Curvelet Transform
DL	Deep Learning
DR	Detection Rate
DWT	Discrete Wavelet Transform
EL	Ensemble Learning
EM	Expectation-Maximization
fc	Fully Connected
FCN	Fully Convolutional Network
FNN	Feedforward Neural Network
fps	frames per second
FT	Fourier Transform
GI	Gastrointestinal
GMM	Gaussian Mixture Model
GRF	Gibbs Random Field
GT	Ground Truth
HoG	Histogram of Gradients

iid	identically distributed
IoU	Intersection over Union
LBP	Local Binary Patterns
mAP	mean Average Precision
MaP	Maximum a Posteriori
MI-RCNN	Mask Improved R-CNN
ML	Maximum Likelihood
MLP	Multilayer Perceptron
MRF	Markov Random Field
MS-RCNN	Mask Scoring R-CNN
MRI	Magnetic Resonance Imaging
NN	Neural Network
OGIB	Obscure Gastrointestinal bleeding
PANet	Path Aggregation Network
PSNR	Peek Signal-to-Noise Ration
RoI	Region of Interest
RPN	Region Proposal Network
SB	Small Bowel
SBI	Suspected Blood Indicator
SGD	Stochastic Gradient Descent
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine
SURF	Speeded up robust features
TL	Transfer Learning
TPR	True Positive Rate

USA	United States of America
WCE	Wireless Capsule Endoscopy

Chapter 1. Introduction

The first chapter begins with the motivation of this thesis work, with a brief description of the problem along the objectives that were proposed. The chapter ends presenting the main contributions as well as the organization of this thesis.

1.1. Motivation

Nowadays, gastrointestinal (GI) lesions are one major problem in the developed countries. For example, in the United States of America (USA), more than 15 million visits to the Emergency Room were related to complaints regarding the GI system and from these, 3 million led to hospitalization [1]. When looking specifically to oncology problems, during 2014 more than 250 thousand new GI cancers were diagnosed and 150 thousand deaths, directly related to these problems, were identified in USA [1]. It is also important to notice that when looking only to USA, the health costs related to GI problems were over 135 billion dollars during 2015. When analyzing data from European countries, it is suggested that the incidence and prevalence of GI disorders are generally increasing, with the highest rates reported in eastern European countries. Also, the incidence and prevalence of many gastrointestinal disorders are highest among older people. With an ageing European population, this will also lead to future increases in disease burden across Europe [2].

The lesions that can appear in the GI tract are diverse and numerous, and there is still a lack of efficient and non-invasive diagnosis techniques, particularly when the small bowel (SB) is the organ to investigate. Since this is a large organ and with difficult access, Wireless Capsule Endoscopy (WCE) appeared as an innovative device with the purpose to easily access the whole SB. This system is considered as minimally invasive and does not cause any pain or discomfort to the patient [3], [4]. Unfortunately, because it generates a large number of images for each exam (more than 60.000), the performance of the physicians in the reading of WCE is disappointing, leading to a lower usage of the device in a clinical practice. Also, the available software tools still present poor performances or a low number of studies so they can be globally accepted by the medical community.

It is then clear that the development of Computer Aided Diagnostic (CAD) systems to be included in WCE are of utmost importance for the gastroenterology's community. With these, the lesions could be detected earlier, increasing the changes of patients to improve their health and consequently reducing the economic spent in the GI-related disorders. CAD systems for WCE also give the possibility of physicians to read each WCE faster, leading the Gastroenterology Departments to have physicians with more time available for tasks where in fact they can be of greater help to the patients.

1.2. Objectives

The main objective of this thesis work is the development of a system that has the ability of detecting different types of lesions in the SB, using WCE exams. This was divided into three sub-objectives:

- In a first phase of the work only two separate types of lesions were considered: angioectasias and tumors. The first was chosen because, in the beginning of the thesis, no papers were found in the literature that addressed the detection of these lesions in WCE, so there was a greater interest in the development of detection tool for this lesion. Regarding tumor detection, this was the continuation of a previously initiated work, so a bigger database was already available for testing.
- A second phase was the development of a multi-pathology framework where different types of lesions could be distinguished. This framework was developed

using the instance segmentation method, with a novel method base on PANet network. This development used the KID Dataset 2, which is a public database with different lesions previously classified and manually segmented by experts.

- Finally, the clinical validation is also one of the objectives of this work. In this case, the system to detect angioectasias was applied to real exams acquired from Hospital of Braga and compared with different physicians.

Regarding the technical considerations, several methodologies were developed and implemented to achieve the objectives previously presented:

- analysis of different features and its performance in different scenarios (specifically textural and color features);
- segmentation of images to separate lesion tissue from non-lesion tissue, to improve classification results;
- ensemble Learning (EL) classification system to use the advantage of different classifiers at the same time;
- convolutional Neural Networks (CNN) based classification and segmentation systems, with the objective of finding not only the localization of the lesion, but also the nature of the lesion.

1.3. Contributions of the work

Several contributions were made during the work of this thesis. Below it is included a list of publications written during the Thesis work (Journal publications, Conference Proceedings and medical abstracts); a list of publications that were submitted but not yet accepted at the time; a list of the two awards won from a part of this thesis work and also the co-supervision of one Master student.

1.3.1. Publications

- D. Costa; **P. Vieira**; C. Pinto; B. Arroja; T. Leal; S. Mendes; R. Gonçalves; C. Lima; C. Rolanda. "Clinical Performance of New Software to Automatically Detect Angioectasias in Small Bowel Capsule Endoscopy". GE - Portuguese Journal of Gastroenterology, 2020. doi: 10.1159/000510024

- D. Costa; **P. Vieira**; C. Pinto; B. Arroja; T. Leal; S. Mendes; C. Lima; R. Gonçalves; C. Rolanda. "Clinical performance of a new software to automatically detect Angioectasias in Small Bowel Capsule Endoscopy". Abstract published in "UEG Week 2019 Oral Presentations". United European Gastroenterology Journal, vol. 7, no. 8 suppl, oct 2019. doi: 10.1177/2050640619854670
- **P.M. Vieira**, N.R. Freitas, J. Valente, I.F. Vaz, C. Rolanda, C.S. Lima. "Automatic Detection of Small Bowel Tumors in Wireless Capsule Endoscopy Images Using Ensemble Learning". Medical Physics, jul 2019. doi: 10.1002/mp.13709
- D. Costa; **P. Vieira**; C. Pinto; T. Leal; S. Mendes; B. Arroja; C. Lima; C. Rolanda; R. Gonçalves. "Performance clínica de um novo Software para Detetar automaticamente angiectasias na endoscopia por cápsula". Abstract presented in "Semana Digestiva 2019" in Poster Session
- **P.M. Vieira**, C.P. Silva, D. Costa, I.F. Vaz, C. Rolanda, C.S. Lima. "Automatic Segmentation and Detection of Small Bowel Angioectasias in WCE Images". Annals of Biomedical Engineering, vol. 47, no. 6, mar 2019. doi: 10.1007/s10439-019-02248-7
- N.A. de Sousa Morais, N. Freitas, A. Cordeiro, J. Torres, S. Anacleto, **P. Vieira**, C. Lima, E. Lima. "A system to automatically detect tumors in white light cystoscopy images using texture". European Urology Supplements, vol. 17, no. 4. may 2018. doi: https://doi.org/10.1016/S1569-9056(18)32476-X
- N.R. Freitas, **P.M. Vieira**, E. Lima, C.S. Lima. "Automatic T1 bladder tumor detection by using wavelet analysis in cystoscopy images". Physics in Medicine & Biology. vol. 63, no. 3. feb 2018. doi: 10.1088/1361-6560/aaa3af
- J. Valente, **P.M. Vieira**, C. Couto, C.S. Lima. "Brain extraction in partial volumes T2*@ 7T by using a quasi-anatomic segmentation with bias field correction". Journal of Neuroscience Methods. vol. 295. pages 129-138. feb 2018. doi: 10.1016/j.jneumeth.2017.12.006
- N.R. Freitas, **P.M. Vieira**, E. Lima, C.S. Lima. "Using cystoscopy to segment bladder tumors with a multivariate approach in different color spaces". 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). jul 2017. doi: 10.1109/EMBC.2017.8036910

- V.F. Silva, R.M. Barbosa, **P.M. Vieira**, C.S. Lima. "Ensemble learning based classification for BCI applications". 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG). feb 2017. doi: 10.1109/ENBENG.2017.7889483
- N.R. Freitas, **P.M. Vieira**, E. Lima, C.S. Lima. "Segmentation of bladder tumors in cystoscopy images using a MAP approach in different color spaces". 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG). feb 2017. doi: 10.1109/ENBENG.2017.7889429
- **P.M. Vieira**, J.F. Ferreira, P.R. Gomes, C.S. Lima. "An adapted double threshold protocol for spastic muscles". 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pages 3630-3633. aug 2016. doi: 10.1109/EMBC.2016.7591514
- **P.M. Vieira**, B. Gonçalves, C.R. Gonçalves, CS Lima. "Segmentation of angiodysplasia lesions in WCE images using a MAP approach with Markov Random Fields". 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pages 1184-1187. aug 2016. doi: 10.1109/EMBC.2016.7590916
- T. Cunha, **P.M. Vieira**, K. Costa, C.P. Santos. "Looking for motor synergies in Darwin-OP biped robot". 2016 IEEE International Conference on Robotics and Automation (ICRA). pages 1776-1781. may 2016. doi: 10.1109/ICRA.2016.7487322
- **P.M. Vieira**, J. Ramos, C.S. Lima. "Automatic detection of small bowel tumors in endoscopic capsule images by ROI selection based on discarded lightness information". 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pages 3025-3028. aug 2015. doi: 10.1109/EMBC.2015.7319029

1.3.2. Publications under evaluation

- **P.M. Vieira**, N.R. Freitas, V.B. Lima, D. Costa, C. Rolanda, C.S. Lima (2020) Multi-pathology Detection and Lesion Localization in WCE Videos by Using the Instance Segmentation Approach. Artificial Intelligence in Medicine **(submitted)**

- N.R. Freitas, **P.M. Vieira**, I.F. Vaz, C.S. Lima (2020) Stochastic Optimization by Using Higher-Order Moments. IEEE Transactions on Pattern Analysis and Machine Intelligence **(submitted)**
- N.R. Freitas, **P.M. Vieira**, A. Cordeiro, N. Morais, J. Torres, S. Anacleto, V.B. Lima, C. Tinoco, M.P. Laguna, E. Lima, C.S. Lima (2020) Detection of Bladder Cancer with Feature Fusion, Transfer Learning and CapsNets. Expert Systems with Applications **(submitted)**
- V.B. Lima, **P.M. Vieira**, N.R. Freitas, C. Rolanda, C.S. Lima (2020) High diversity-based and brain-inspired mixture of experts classifier for automated ulcer detection in wireless capsule endoscopy images. Pattern Recognition **(submitted)**

1.3.3. Awards

- National Scholar Award 2019 – UEG (United European Gastroenterology) Week 2019
- Travel Grant for Best Submitted Clinical Science Abstracts - UEG (United European Gastroenterology) Week 2019
- "Prémio Nacional da Gastrenterologia 2019" - Sociedade Portuguesa de Gastrenterologia

1.3.4. Supervision

- Catarina Pinto Silva. "CMEMS-Uminho Algorithm: a novel method for Automatic Detection of Angioectasias in Small Bowel Capsule Endoscopy". School of Medicine – University of Minho. December 2019

1.4. Thesis' Organization

This document is divided into ten chapters, where the first and current Chapter was the introductory section with a brief overview of the main objectives and contributions of the work; and the last Chapter is the Conclusions section, with also an overview of the future work.

Chapter 2 presents a context of the clinical aspects of this work, with a special focus on the GI system and the main lesions that are found on it. Also, an explanation of how WCE works and its main advantages and disadvantages are presented. Finally, an overview of the available software for WCE reading and how physicians perform when using that software.

Chapter 3 is the section where the state-of-the-art of methods for WCE analysis will be presented. In this, different methods for reducing video duration, image quality enhancement and lesion detection will be explained and their results analyzed.

Chapter 4 consists in the presentation of the methods that were used in the experimental work of this thesis. This includes the color analysis of the lesions, features acquisition and segmentation, and also all the classification methods that were used in this work.

Chapter 6 to Chapter 9 covers the results of this work, separated according to the different topics of this Thesis. Chapter 6 is based on two different publications [5], [6] and addresses angioectasia detection in WCE images. Chapter 7 presents results from the tumor detection and is based on one publication [7]. Chapter 8 shows the results from the multi-pathology detection section, using CNNs and instance segmentation for segmentation and classification of lesions, being based on [8]. Finally, Chapter 9 shows the results of the Clinical Validation of the angioectasia developed software within the Hospital of Braga, and was published in [9].

The last chapter is dedicated for the main conclusions taken out from this thesis, where the different contributions for the field will be shown. Also, it is analyzed the future work around medical imaging techniques applied to Capsule Endoscopy exams.

Chapter 2. Clinical Context

This chapter begins with a brief overview of the GI system and the lesions most commonly found in it. After that, the WCE will be presented, along with its main advantages and disadvantages. In the end, the commercially available software will be shown, along with some studies about the performance of physicians related to WCE readings.

2.1. Gastrointestinal System

The digestive system consists of a number of hollow organs, from the mouth to the anus [10]. The upper part of the GI tract includes the mouth, pharynx, esophagus and the stomach; while the lower part consists in the SB (that can be divided in duodenum, jejunum and ileum) and the large bowel (cecum, colon, rectum and anus) [11].

The whole GI tract can reach more than 9 meters, and the SB is the longest length organ in this system. In fact, the average length of the SB is of 6.9 meters for males and 7.1 meters for females [12]. Due to its extension, the transportation of the food (called chyme when in the SB) is done with a specific mechanism called peristalsis. This process consists of a series of radially

symmetrical contraction and relaxation of muscles, which help the propagation of the chyme [10]. This mechanism is also the responsible for transporting WCE along its way.

2.1.1. Gastrointestinal Lesions

There are a variety of lesions that can appear in the GI tract, but in this work only the ones that appear more frequently in the SB and that can be visualized in a WCE exam will be analyzed. So, the ones that will be addressed are ulcers, angioectasias, bleeding, polyps and tumors. It is important to notice that bleeding is not in fact a lesion itself, being a situation that can appear due to other lesions. But, because of its nature, a bleeding is clearly visible in a WCE exam, and usually hides the lesion that was responsible for the situation [13].

2.1.1.1. Ulcers

These lesions manifest themselves as erosions of the intestinal wall, and appear more often in the duodenum [14]. Ulcers can be characterized by their number (single or multiple ulcers) and shape (circular or oval). When superficial and affecting only the two first layers of the mucosa, they are specific types of ulcers called erosions. Conventional ulcers are usually characterized as lesions extending into a deeper layer of the underlying tissue [15].

Ulcers usually appear when tissue is destroyed by gastric juices. These fluids are produced by the intestine to digest the different nutrients in food, like starch, fat, and protein. Basically, an ulcer can turn into three types of complications: bleeding ulcer; perforated ulcer; and narrowing ulcer [16]. A bleeding ulcer occurs when the ulcer erodes one of the blood vessels; which leads to hemorrhage. Perforated ulcers appear as a hole in the wall and can lead to intense abdominal pain. Narrowing ulcers (or ulcerated strictures) cause a stenosis (stricture) of the SB, which can lead to severe vomiting [17]. The prevalence of peptic ulcers (the ones that appear in stomach and duodenum) is described as being in order of 1 case in 1000 person-years in the general population [18].

Figure 1 shows two examples of ulcers taken from a WCE exam (in this case, two narrowing ulcers). Here, they appear with a clear white coloration with a reddish tissue area around them.

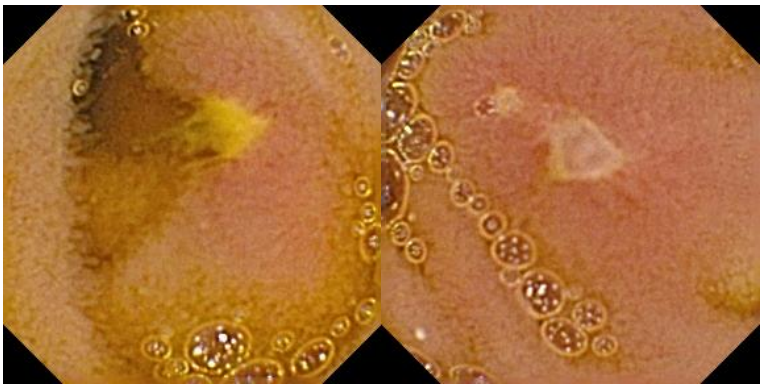


Figure 1. Examples of ulcers taken from a WCE exam.

2.1.1.2. Angioectasias

Angioectasias, sometimes referred as angiodysplasias, are degenerative lesions of previous healthy blood vessels, which are caused by microvascular abnormalities that may appear in the mucosa or submucosa of the SB wall [19]. These lesions are the cause of approximately 8% of all GI bleeding episodes, being the most common source of bleeding from the SB in patients older than 50 years [20], [21]. Angioectasias are the most common vascular malformation of the GI tract in the general population. Most of these lesions are detected in patients older than 60 years. Angioectasias may be asymptomatic or may present signs and symptoms consistent with GI bleeding [22].

Angioectasias have a cherry red appearance, with a diameter from 2 to 10 mm. They are superficial lesions, therefore easily spotted by imaging techniques that capture images from the inside of the GI tract [23]. Three examples of these lesions can be seen in Figure 2.

Figure 2. Examples of angioectasias taken from WCE exams.

2.1.1.3. Bleeding

As stated previously, a bleeding episode (or a hemorrhage) is not effectively a lesion, being however a consequence of it. Therefore, it is of great importance its detection for diagnosis purposes. It is estimated that upper GI bleeding, lower GI bleeding and obscure GI bleeding (OGIB) account, respectively, for 50%, 40% and 10% of total GI bleeding [24].

OGIB situations are those where standard evaluation fails to reveal the source of the bleeding [25]. In these cases, the localization of the source of the bleeding is of utmost importance for the stabilization of patients. Also, it is important to notice that the majority of OGIB-associated lesions are present within the SB, an area which of difficult access for conventional diagnosis methods [26].

Figure 3 shows three examples of bleeding situations, and where the heterogeneous nature of this episode is evident.

Figure 3. Examples of bleeding in images from WCE exams.

2.1.1.4. Polyps

Polyps appear due to an abnormal growth of tissue of the intestinal wall. SB polyps present variable shapes and usually the symptoms are slow to appear, causing a delay in diagnosis by various months [27]. These symptoms are present in 40-70% of patients and can include abdominal pain, iron deficiency anemia or occult blood loss, weight loss, nausea and vomiting, and intermittent obstruction [28].

Polyps can be divided in two different categories: non-neoplastic and neoplastic. Non-neoplastic polyps include hyperplastic polyps, inflammatory polyps and hamartomatous polyps [29]. Usually this type of polyps does not carry a risk of developing into cancer. When looking at

the neoplastic polyps, which include both adenomas and serrated polyps, the probability of leading to cancer is higher. Another important fact is that larger polyps have a greater risk of cancer, especially with neoplastic polyps [30]. Some examples of these lesions can be seen in Figure 4.

Figure 4. Examples of polyps taken from WCE exam.

2.1.1.5. Tumors

A tumor, as a general term, refers to a solid or fluid-filled cystic lesion, which might or might not have formed due to an abnormal growth of neoplastic cells. Tumors are also referred to as those neoplasms that often form an abnormal mass of tissue.

When faced with OGIB in patients under the age of 40, the most frequently detected lesion are SB tumors [31]. However, SB tumors are considerably rare when compared to other GI tumors, since they are just 3-6% of the prevalence of these lesions [32].

The clinical presentation is nonspecific and vague, since symptoms can be of different natures, including anemia, gastro-intestinal bleeding, abdominal pain or SB obstruction [32]. The nonspecific nature of these tumors' symptoms exist also because there are several types of SB tumors. The most frequent ones are Adenocarcinoma, Lymphoma; Carcinoid and Gastrointestinal stromal tumors [33]. In Figure 5 some examples of tumors taken from WCE exams can be seen.

Figure 5. Examples of tumors taken from WCE exams.

2.2. Wireless Capsule Endoscopy

Conventional endoscopy only reaches the duodenum and the colonoscopy only reaches the ileum, so when using these techniques, the largest portion of the SB is "invisible". The double-balloon enteroscopy was the first technique that allowed the entire visualization of the SB, and consisted of the use of a balloon at the end of a special enteroscope camera and an overtube that is put inside an endoscope. Although allowing a real-time visualization of the whole GI tract, the procedure can take more than 3 hours and is of great invasiveness for the patients [34]. Other techniques are Computed Tomography (CT) scan and Magnetic Resonance Imaging (MRI), which allow the visualization of the small bowel but do not allow an analysis of the internal tissue, and the patients are exposed to radiations (in the case of the CT).

2.2.1. The Device

WCE is a medical device that was introduced as a novel technology that, contrary to the conventional endoscopy/colonoscopy, allows the inspection of the entire GI tract without major risks and discomfort to the patients, not requiring specialized endoscopic operators [35].

This device began its development in 1981 by Gavriel Iddan; but due to the low-level technology of the time, two decades were needed for this product to be available. So, only in 2001 it was fully developed by the company Given Imaging [36], and approved by both the European Commission and the Food and Drug Administration. The first capsule was named M2A™, where its appearance and its components can be seen in Figure 6.

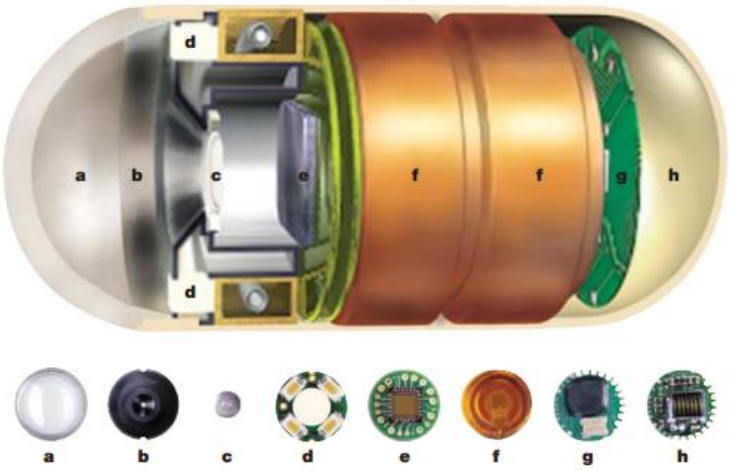


Figure 6. Representation of M2A™, the first generation of WCE (top) and all the components in the device (bottom). Adapted from [4].

Although since 2001 several similar systems from different manufacturers reached the market, their operation is quite similar. These pill-like devices include a miniaturized camera, a light source consisting usually on four LED's and a wireless circuit for the acquisition and posterior transmission of signals. After the acquisition, the video frames are wirelessly transmitted to an external receiver, worn in a belt by the patient, and stored in a hard drive. The images are captured by a short focal length lens, as the capsule is propelled through the GI tract [35].

Since M2A™, afterwards renamed as PillCam™ SB, several WCE models appeared in the market. The company Given Imaging, that was later acquired by Medtronic, developed two more generations of WCEs for the SB (PillCam™ SB2 and PillCam™ SB3), a version specially tailored for the esophagus (PillCam™ ESO) and one WCE developed to visualize the large intestine (PillCam™ COLON) [37]. Companies like Olympus (with Endocapsule), IntroMedic (with MiroCam), among others, developed their own WCE systems, each one with specific characteristics. From all the specifications, the major differences are in the frame rate, angle of view and frame resolution.

The first system (PillCam™ SB), had a frame rate of 2 frames per second (fps), a frame resolution of 256 x 256 and an angle of view of 140° [3]. In Table 1, some specifications of 5 different models of WCE systems were grouped. As it can be seen, both resolution and frame rate present disappointing values, especially when compared to conventional endoscopes that have a video feed of 30 fps and some of them present images with Full HD resolution [38]. Another important point is the difference of values of CapsoCam SV1 when compared to the remaining

devices. This happens because this specific WCE has several cameras around its entire body, being able to acquire different images at the same time which can be afterwards registered along side to show a 360º view of the GI tract. This is also the reason why the frame resolution is given as the number of pixels and not has the product between width and height. Although not shown in this table, it is important to notice that the prices are similar to every capsule except OMOM capsule that costs almost half the price of the remaining models [31].

Table 1. Comparison of the specifications of different models of WCE in the market. Information taken from [31], [39]–[41]

Model	PillCam™ SB3	Endocapsule 10	CapsoCam® SV1	MiroCam®	OMOM
Year	2013	2016	2011	2012	2005
Frame rate (fps)	2-6	2	12-20	3	2
Angle of view (º)	156	160	360	170	140
Frame Resolution	320x320	512x512	221884 pixels	320x320	640x480
Battery time (hours)	12	12	15	12	8

As it can be seen in Table 1, the resulting video has a duration of several hours and consists of thousands of images (usually more than 60 000 frames per exam). The analysis of this large number of images can be a boring task for a physician, and consequently, is predisposed to subjective errors since most frames contain only normal tissue [42].

2.2.2. Software and Physicians' Performance

Each WCE model described in Table 1 has also its own software to help physicians to analyze the exams. In the different software there are currently two tools with the aim of detecting specific problems in WCE exams: "Suspected Blood Indicator" (SBI) and "Top 100", both found in RAPID software used in PillCam™ capsules.

The SBI tool has the purpose of detecting frames in the presence of blood and red areas [43] and was introduced in the second version of capsules [44]. During the last years, several published studies analyzed SBI performance of detecting the presence of blood and several lesions, including angiodysplasias. A meta-analysis review for the validity of the SBI in clinical context clearly shows the high sensitivity of this tool for detecting active bleeding (around 99%), although it still shows a low specificity (around 65%). Yet, the detection of lesions with bleeding potential is clearly reduced, with values of 55% and 58% for sensitivity and specificity, respectively [45]. When looking for angioectasia detection only, a significant decrease is shown, with the different works showing values of 13% and 85% [43]; 41% and 68% [46] and 39% and 54% [47] of sensitivity and specificity, respectively. When looking at these values, it is clear the limitation of using SBI when lesions with bleeding potential, specifically angioectasias are present.

"Top 100" tool was included more recently in RAPID© software (end of 2016) and its purpose is to show a preview of the one hundred images that most likely include an ulcer, bleeding, or a polyp [48]. In the Software User Manual is stated that this feature "is not intended to detect pathology in lieu of a physician and should not be used as a substitute for reviewing the entire video" [48]. Two studies appeared recently to analyze this tool in the clinical practice. In the first one [49], 81 exams were included, and "Top 100" correctly identified 26/30 (86.7%) patients with angioectasias and 27/34 (79.4%) patients with ulcers. Also, active bleeding was present in 4 patients, and all were diagnosed by "Top 100". For overall detection of significant lesions, 164/215 (76.3%) of lesions were identified [49]. In the second study [50], 71 patients were included, where "Top 100" detected 64/67 (95.5%) of angioectasias, 17/30 (56.7%) of ulcers and all instances of active bleeding (9/9). Also, all the false positive images that "Top 100" selected were easily identified and promptly ignored by the experienced reader [50].

The various software have different viewing options. One of them is the option to analyze more than one image at the same time. Looking at Figure 7, in (a) one image is seen at each time, but in (b) and (c) two images are seen at the same time. It is possible also to visualize four images at the same time in some of the software tools, increasing the speed of visualization (from 5 to 40 fps) [51]. Both increasing the number of images in the screen and increase the speed of visualization will decrease the time needed for evaluation of the exam (which can take from 30 to 90 minutes), but have also the ability of worsening the performance of the physicians [52].

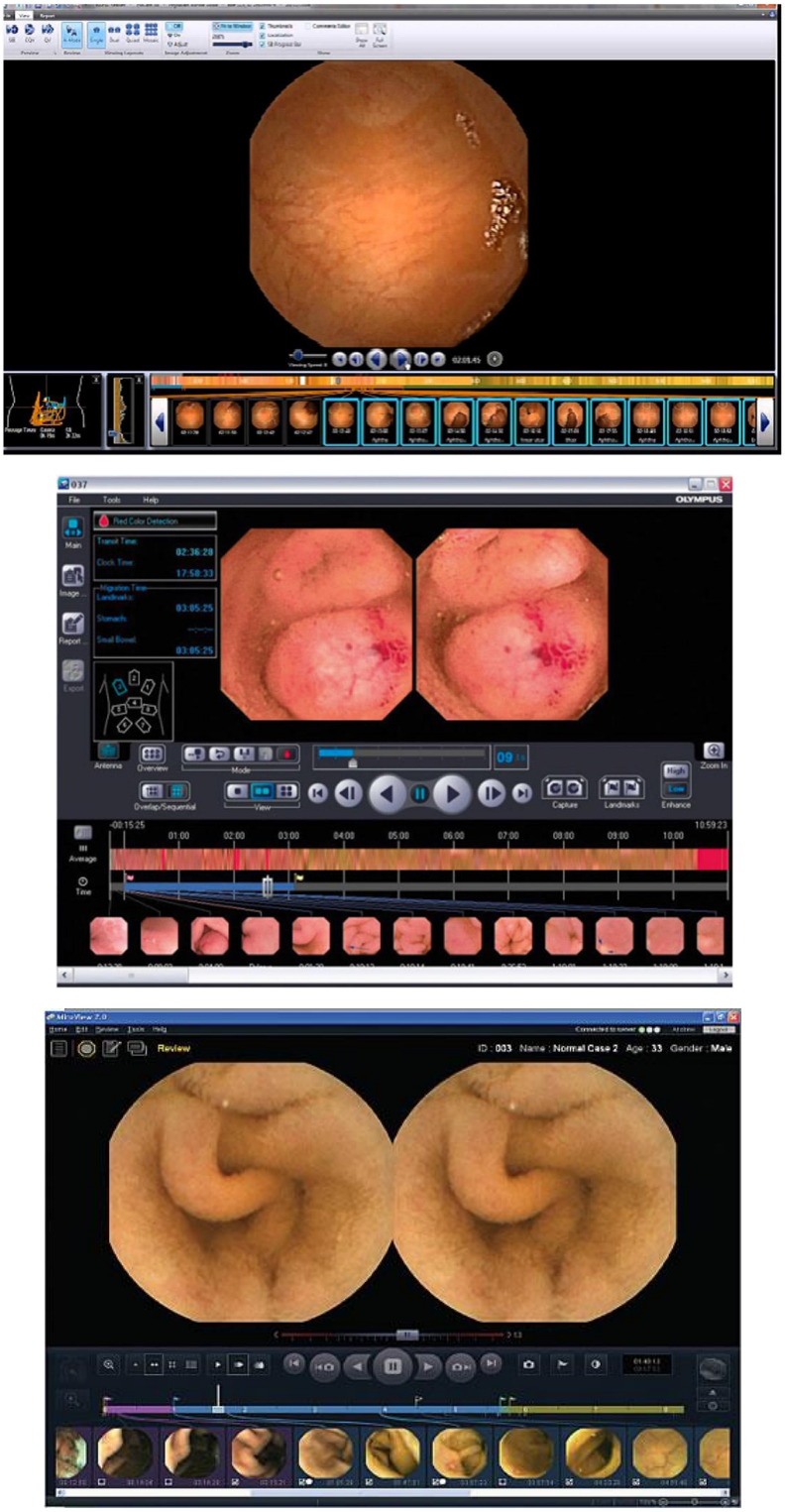


Figure 7. Screenshots of different WCE models' software. (a) RAPID Reader from PillCam SB (b) EndoCapsule Software (c) MiroView from MiroCam. Adapted from [53], [54]

It is important to notice that the performance of physicians in the analysis of WCE exams is rather low, when compared to other methods of diagnosis. In a study of 2012, 17 WCE readers (8 with low experience and 9 with median or high experience) were analyzed [52]. Although the

performance of physicians when viewing entire WCE videos was not evaluated, since 24 small clips were the object of study; it is possible to infer some conclusions about this specific subject. It is rather shocking that in average, less than 50% of the lesions were found by this group of physicians; with a range that goes from 17% to 78%. Also, it is interesting to notice that there was no direct relationship between the reader experience in WCE analysis and the performance in this study (the physician with a performance of 17% was in fact the one with most experience in WCE). Regarding the different lesions in the clips, angioectasias were the lesions most found (69%), followed by polyps (46%), ulcers (38%) and blood (17%).

2.3. Conclusion

In this chapter, an overview of the GI system was presented along with the main lesions found in the SB (bleeding, angioectasias, ulcers, tumors and polyps); which have different visual characteristics that can be used for their detection. The device known as WCE was also described in this chapter, along with its main advantages and characteristics. Along with these, the differences between the main manufacturers and the existing software were presented. In the end, an analysis on the performance of physicians and software within the scope of lesions detection in WCE exams is done.

Chapter 3. Computer Aided Systems for WCE: An Overview

This chapter presents a brief overview of some of the developed methods that improve the analysis of WCE exams. These methods are divided into three different sections. While the first covers the methods directly aiming to reduce the visualization time, the second refers to methods that increase the quality of the images/videos and in the final one, methods for automatically detect lesions will be addressed.

3.1. Reduction of Reading Time

Reduction of the time needed to read a WCE exam is extremely important, since it is an extremely time-consuming tasks for the physicians, as was stated in the previous chapter. WCE exams have a huge quantity of data, becoming difficult for these professionals to have their entire attention for reading the images. In the last decade several works have appeared with the purpose of reducing the size of the videos, so a better and fast analysis is possible.

To accomplish this objective, one of the most used techniques is data mining, whose purpose is to find implicit information in a big dataset. Data Mining methods, rather than trying to find specific lesions in the tissue, search for frames with content that deviates from the normal behavior (they use the knowledge that there are a big percentage of frames in each video that does not have any lesion or anomaly). One of the works that reduces the video size of WCE exams considers the set of images as members of a spatial vector, using then Non-Negative Matrix Factorization algorithms [55], [56]. This method is used in subsets of consecutive frames where a number of the most representative orthogonal frames are chosen. These were able to reduce the number of frames to 85% of its original size, not compromising the content of the video. When a larger database was used, a method based in the same algorithm reached a reduction of only 66% [57], leading to the conclusion that this method in clinical practice may not be very efficient. Another solution that sometimes is used in this area is to remove frames that only presents intestinal content that has not significant information for lesion detection or diagnosis. Examples of this intestinal content can be bubbles, fluids, intestinal residues, food remains, among others. Frames that contain this type of content can be automatically detected through algorithms, similar to those used to detect lesions and pathologies, that are specifically designed with the purpose of finding certain characteristics of these structures. In [58], non-informative frames with this specific content were found using Local Binary Patterns and K-Nearest Neighbours, having described a detection rate of 99% of these frames.

Other methods that are currently used to reduce the volume of images in a WCE exam is the removal of redundant frames. These frames appear in every video, being usually more than 30% of the total number of frames [59]. They appear because the WCE device, due to the peristalsis movements, present different speeds and it stops sometimes. Since the frame rate is constant, when WCE moves slowly or is stopped, consecutive frames will present the same region of tissue [59]. This process can be made analyzing features from consecutive frames. These features can be of different natures: color, texture or even both. In [60], MPEG-7 standard features are used as visual descriptors of WCE frames. After, the difference between consecutive features is computed so similar frames can be found. The underlying thought is that, considering a threshold applied to those differences, the selected frames will be frames with certain characteristics that are significative different from the majority. This method reached values of precision around 70% with four clips of 200 images each. Another way for removing redundant frames is by estimation of WCE movement inside the GI system, existing two different methods to reach this objective. The

first uses the SURF (Speeded up robust features) or SIFT (Scale-Invariant Feature Transform) algorithms, to look for salient points in each image; which are found with the computation of hessian matrices [61]. These points are then compared in consecutive frames so they can be matched. This way, a transformation which matches these same points in consecutive images is obtained, which is then used to find similarities among frames. The second method computes a model that estimates the movement of the WCE inside the GI system [62]. With this, the speed of the movement can be also estimated, and if it is below a certain threshold it means that the frame that is immediately after contains the same tissue and consequently can be removed. This method uses two different modules for constructing the model, one of them is rigid (consisting in the movement of the device) while the other is non-rigid (being related to the natural peristalsis movement of the small bowel). While the first method [61] reached reductions of 52% and was tested with a video of 30 000 frames; the second method [62] reached an average reduction of 68% with several clips of 100 frames. Another approach applied the reduction of both non-informative and redundant frames; applying a Gauss-Laguerre Transform and a distance measure based on image edges (using the Canny operator) to reduce these two types of frames, respectively [63]. This method reached an average reduction of more than 50% of frames without any loss of useful frames and was applied to six full exams of WCE, which is a very promising result.

3.2. Improvement of Image/Video Quality

One of the main problems of WCE exams is the low quality of the optical instrumentation and the illumination of the available systems. Since this device must have an autonomy of at least 8 hours, and its size cannot be higher than already is, there is a strong limited capacity for energy consumption. In this way, WCE cannot have the most appropriate image acquisition specifications to correctly represent the inside of the GI system. For example, the quality of camera and illumination in conventional endoscopy is significantly better when compared to WCE, specially because there is no need for energy autonomy for these devices. Because the GI system is very dark, a lack of illumination will lead to images where some of the structures are not easily identified. On the contrary, and because of light inhomogeneity, there is also problems with light saturation when some structures are near the camera [64]. One of the greatest difficulties when analyzing these methods is the lack of metrics to evaluate them. In this way, the only way to compare the different methods is a visual verification, which is subjective.

The first chosen method is the use of the adaptive contrast diffusion to improve and enhance images [65]. This uses an anisotropic filter into the images, which changes their contrast without amplifying the present noise. Although stating and presenting promising results, only experiments with images with darker zones were shown, so the improvement of images with saturated zones would be necessary. Another approach is to use homomorphic filters to improve image quality [66]. In this work, Fourier and Wavelet transforms based Homomorphic filters were used in the channel Y of the YCbCr color space. With this, low frequency components are removed, with the purpose of reducing the significance of the lighting that shines on the tissue due to the light source. This method seems to present better results than the previous ones, but apparently they artificially increase the clarity of the images, which in the case where saturation of the tissue appears, this method will probably fail.

Another way to improve the images is by creating a new color space, specially designed for these specific images, which is the case of Vu et al. with the GI Color Space [67]. This new color space is RGB-based and highlights the most prominent characteristics by using a learning scheme based on a self-organizing map to handle the GI color components. Afterwards, an histogram equalization is performed and the new image added to the new original one, showing and enhanced region of interest (RoI). One the main limitations of this method is the necessity of a physician to choose the RoI to enhance in the image, which reduces its applicability in the clinical practice.

There are also interesting methods that were developed to enhance images taken from conventional endoscopy/colonoscopy. Although they were not used in WCE images, their findings are important also for this field. One of the examples in the use of the conventional algorithm Retinex, which improves the lighting conditions [68]. This method has the ability of modelling the lightness and color perception of human visual system in images and in the resulted images shown in the paper, the results seem promising. In [69], an adaptive sigmoid function is applied to Y component in YCbCr color channel with a posterior space-variant color reproduction method. Although it is stated that this method can improve lesion detection in these images, it is unclear if the improvements really can work in WCE since the resulting images seem to have a too high contrast. Finally, [70] developed a multi-step algorithm that applies the TV model to obtain the cartoon-texture decomposition of the input image. Afterwards, a Texture-Enhanced Histogram Equalization is made to the cartoon-texture components to get the enhanced images. Although

presenting a reduction of the Peek Signal-to-Noise Ratio (PSNR), it is difficult to understand the visual results in the provided examples. These three works could be of inspiration for use in WCE images, but as the resolution, illumination and field of view of these acquisitions are significantly different from those of the endoscopy, these methods should be used with caution.

3.3. Automatic Detection of Lesions

Automatic detection of lesions is the area where more works have been appearing in the last decade. The lesions where more studies are focused on are bleeding, Chron's disease, polyps, tumors and ulcers. More recently, several works have been covering the detection of angioectasias (or angiodysplasias), since the medical community has shown an increase of attention given to these lesions and the future problems they can cause to the patients.

To analyze the evolution of the results from the automatic lesion detection in WCE, Figure 8 was created. In this, it is shown the performance variation of the existing methods over the years (in blue, from 2000 to 2009, and in orange, from 2010 to 2014). The values in this figure were taken from two review articles ([71], [72]). It is noticeable the absence of Chron's Disease methods during the period of 2010 to 2014, which happens since there was not a significant number of paper in this area. The lack of works in this area may have happened because no improvements could be achieved. Angioectasia methods are also not present in Figure 8 because the first work regarding the detection of this lesion only appeared in 2016. Another interesting point for analysis is the variation in the several types of lesions. It would be expected that with the technological advances and the improvement in several techniques (like machine learning methods), improvements in the performance of the detection of most pat of the lesions had happened. But, in fact, this not happened for tumors, perhaps due to the high variability of these lesions, where several different types can be found, and consequently the used datasets can affect in a great manner the achieved results.

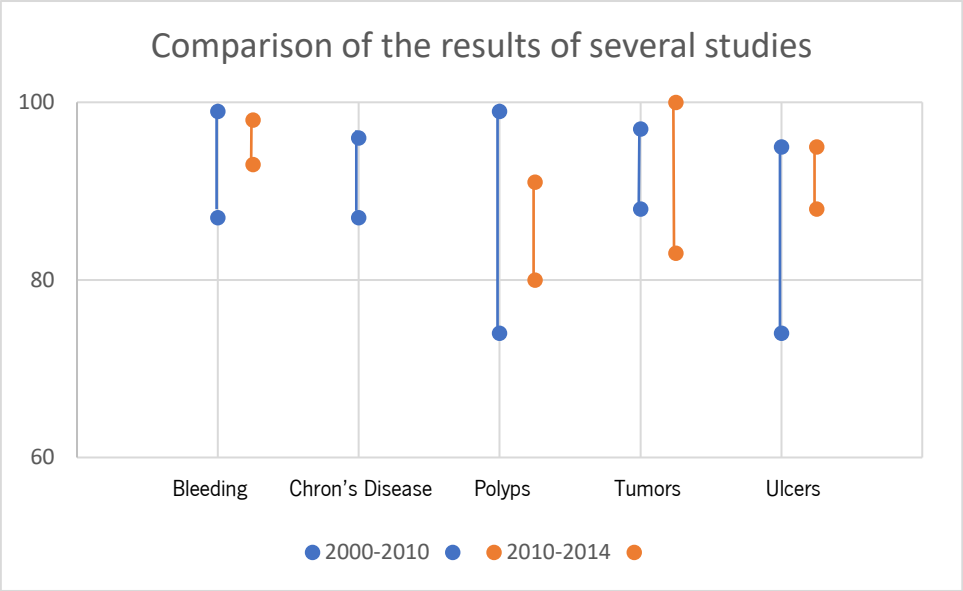


Figure 8. Evolution of the results of studies with the purpose of detecting several lesions in WCE exams. Minimum and maximum performances from 2000 to 2009 in blue, and minimum and maximum performances from 2010 to 2014 in orange.

Since 2014 a lot of new methods have appeared with performance improvements, specifically due to the use of deep learning based computational algorithms.

The following sub-sections will address some of the methods developed to detect each of the lesions, with the last one presenting methods designed to detect multiple lesions at the same time.

3.3.1. Bleeding Detection

The first works regarding lesion detection in WCE systems addressed bleeding detection, since this is the most easily distinguishable among the lesions in the SB. These have usually a homogeneous appearance, with a color that is very distinguishable within the tissue and when comparing with other structures that appear in the videos.

The first study with the purpose of detecting blood in WCE frames proposed a segmentation of bleeding regions, using the Expectation-Maximization (EM) algorithm for clustering and the Bayesian information criterion for helping to choose the number of clusters, using RGB color channels as observations [73]. A frame is considered as bleeding if after the segmentation, a blood region is found. In the same year, a method was developed where images were converted to the HSV color space and afterwards, with a rigid division of the images the intensity of the resulting blocks were compared to pre-defined thresholds [74]. This first method reported a sensitivity of

88%, using 5 sets of 400 images each. A study of 2008 presented a method which used features both from cooccurrence matrices between the different channels and histogram measures from the dominant colors to be an input to an ensemble of SVMs [75]. This method achieved a sensitivity of 82% when using a dataset of 550 bleeding images. Still regarding bleeding detection in WCE exams, a new method that can adapt to each patient was proposed in [76]. This method uses the color spectrum transformation method with a new parameter compensation step which balances irregular image condition based on the color balance index. With this compensation step, a sensitivity of 95% was achieved (compared with an 80% without the compensation step), in a dataset of 4800 images. This work is important to understand the effect that the variability among patients and capsules causes to the images and consequently to the detection methods. There are also works that detect bleeding lesions with machine learning techniques. For example in [77], a probabilistic neural network classifier was used to detect blood in pixels using intensity values from both RGB and HSI color spaces. This is a method that classifies each pixel independently on the image where it is, so at a pixel level, a sensitivity of 88.6% was obtained (using 20 bleeding images). At an image level (using 15000 images), a sensitivity of 93% and a specificity of 86% were obtained; which is a very promising result, taking into account the size of the used dataset. Also using machine learning techniques, in [78] an SVM was used to classify bleeding images using Pyramid of Color Invariant Histogram to extract features. With this technique, histograms of both HSI and RGB were computed in several rigid blocks of the image, and then divided in a pyramidal model. In a dataset of 560 images, an accuracy of 98% was reached. More recently, deep learning was used in segmentation of blood regions in WCE images [79]. This specific work compares the use of a simplified MLP and a simplified CNN for the classification of each pixel as being blood or tissue. The DICE score was higher in the CNN model as was expected (87%) but was only 4% higher than the MLP model (83%), which is a good result, having into account the reduction of parameters in the MLP model.

3.3.2. Angioectasia Detection

In spite of the high number of papers involving WCE research only a few refer specifically to angioectasias. A saliency maps based approach from RGB color space (only channels red and green) was proposed in [80], [81], where a sensitivity of 94% and a specificity of 84% were reported in a database with more than 3600 images. In [82], a histogram equalization step is used to increase image contrast with a posterior decorrelation between RGB channels to enhance color

differences in the images. To select a RoI, a threshold is used in the green channel to provide a seed to a region growing algorithm. The resultant regions are then splitted when the variance is high enough. Then 24 statistical textural and geometrical features are extracted for several color spaces. A decision tree is used to classify each region as normal and abnormal reaching an accuracy of 96.8%.

Handcraft (color and texture) and DL based features were compared in [83] for angioectasia lesion detection in a database of 600 frames having been reported a sensitivity and specificity of 62% and 78%, respectively. A DL based architecture for pixel-wise segmentation purposes was proposed in [84] by using AlbuNet and TernausNet networks where a Dice coefficient of 85% in a database of 600 images was reported. Although it is stated that it can be used for a detection purpose, the paper does not present these results. Also using DL, in [85] the authors train a CNN system based on Single Shot Multibox Detector using 2200 images. Afterwards, the results are obtained training the system with a dataset of 10500 images (500 of them with angioectasia), reaching a very high sensitivity (98%), but a rather low Positive Predictive Value (75%).

3.3.3. Tumor Detection

Most of the state-of-the-art automatic tumor classification systems for WCE frames rely on texture extraction algorithms since texture changes are usually what the physicians search for [86], [87]. Abyoto et al. [88] stated that the textural information is localized in the middle and higher frequencies of the original signal. Szczypiǹski et al. [89] proposed a model of deformable rings to compute motion-descriptive characteristics, however only a 2% reduction of time spent into reviewing the video was achieved although a significant 20% increase of detected lesions by 10 expert physicians was reported. In [90]–[92], discrete wavelet transform (DWT) was used to select the bands with the most significant texture information for classification of tumors in WCE exams. Furthermore, in Barbosa et al. [93], the measurement of the non-gaussianity of these statistical texture descriptors were used, in a classification scheme to identify abnormal frames. Co-occurrence matrices and Haralick texture descriptors were also used in a more recent study[94], where mutual information was applied to select the most relevant features and an Adaptive Neuro-Fuzzy Interface System was developed to classify each image regarding the presence of tumor, reaching an accuracy of 95%. Szczypiǹski et al. [95] developed a framework based on Sequential

Forward Floating Selection (SFFS) for feature selection and Support Vector Machines (SVM) for classification with the purpose of detecting several types of lesions. Also using SVMs, Liu et al [96] developed new texture features, combining discrete curvelet transform with fractal technology; which were reported as being robust to illumination variation of the images. The main drawback of some of the approaches described before is the computational resources required for DWT, Inverse DWT and co-occurrence computations[97]. Although some new implementation strategies for co-occurrence matrices have recently appeared [98], [99], there is usually the need for new hardware for them to become competitive, especially when a high number of frames need to be computed (which is the case of WCE exams).

3.3.4. Polyps Detection

Regarding polyps detection, the main characteristics used for the detection are their distinctive shapes and also their color and texture [100]. The first work regarding this subject appeared in 2007 and used an approach for texture characterization based on the chromatic and achromatic domains of the images [35]. Afterwards, a texture unit is computed that characterizes the local texture information for a given pixel and its neighborhood, and then a Radial Basis Function is used for classification. A 96% accuracy is reached, but the dataset only contained 70 images for testing (35 of them containing a polyp). In another work, two different shape features were used for discriminating polyps from normal regions [101]. The angular radial transform (an MPEG-7 shape descriptor) and Zernike moments are used as features, along with an MLP neural network as the classifier. This approach was tested on 300 images (150 of them contained polyps), achieving an accuracy of 86%. In this, only shape information was used, having discarded both color and texture information. In a related work by the same authors, the combination of color (HIS color space) and shape features was exploited [102]. A compressed chromaticity histogram (hue and saturation) and Zernike moments were the chosen features to be classified by MLP. With the same database used previously, an accuracy of 94.20% was reached, leading to the conclusion that color in fact is an important feature for polyps characterization. A different method used Log–Gabor filter based segmentation along with an edge detector (SUSAN) an active countour segmentation to find polyps in WCE images [103]. This reached a sensitivity of 100%, but it is important to notice that a dataset of 50 images was used (10 of them with polyps). Later the same authors extended this work using an SVM classifier with the same dataset [104].

In [105], LBP texture features are computed over contourlet transformed images with an SVM classifier, reaching 97% of accuracy. Unfortunately, the total number of frames used for test this method is not stated so it is difficult to compare it with other works. The same authors later used a combination of color and textural features (total of 118 different features) that was reduced to a number of 13 features using a Laplacian eigenmap method along with a K-Nearest Neighbors classifier [106]. These features were then used as input for a boosted classifier, leading to an accuracy of 90% in a 1200 images dataset. Still the same authors applied an uniform LBP on DWT sub-images to extract features to be used in an SVM [107]. In the same dataset as previously the accuracy increased to a value of 91.6%.

In [108] a protrusion measure based on mean and Gaussian curvature is defined to use as features for polyp detection. This method uses a threshold value for classifying each image, reaching a value of accuracy of 80% in 17 short videos of 100.frames each. This protrusion measure was later evaluated in a dataset of 400 images (200 of them polyps) with an SVM classifier, leading to an accuracy of 65.5% [109]. The same work also tested the dataset combining protrusion measures with basic LBP features, leading to a increase of accuracy to 97%, and when multiscale LBP was used an accuracy of 99% is reached.

More recently, deep learning was also used for polyp recognition. In [110], the authors use a sparse autoencoder with an image manifold constraint, which makes the learned features to preserve large intervariances and small intravariances among images. This model was tested in a dataset of 4000 images, reaching a sensitivity of 98%. A deep CNN was also implemented in the same work, reaching a lower value of sensitivity (97%).

3.3.5. Ulcer Detection

The most usual types of features to classify ulcers are both color and textural features. For example, in [111] was proposed the computation of Chebyshev polynomials along with chromaticity moments in HSI color space. Each image is divided into 36 patches, and each of them is classified independently with an MLP. 3600 patches of images are used (half of them containing ulcers), leading to a sensitivity of 93.0%. The same authors proposed also a curvelet-based LBP applied to channels from YCbCr color space [112]. Using the same database, a similar result was achieved (sensitivity of 93.3%).

In [104], the detection of ulcers is based on the computation of log Gabor filters in HSV color space and applying a segmentation scheme based on a fuzzy region-growing approach. Haralick texture descriptors are then fed to a SVM classifier, reaching a sensitivity of 75% with a dataset of 50 frames (20 with ulcer), which is a rather low result when looking to the size of the dataset. A bag-of-words model, together with feature fusion, is used in [113], where several image patches are used to extract LBP and SIFT features. These are then used as input for a SVM classifier having reached an accuracy of 90% for a dataset with a size of 350 images. In [114], a two-staged detection system to detect ulcers is proposed. In the first stage, an automatic estimation of salient regions across the WCE images is done using a superpixel representation. A saliency map is built using both color and texture information from the superpixels. In the second stage, the obtained saliency map is coded with a modified Locality-constrained Linear Coding method, using different descriptors (SIFT, HoG and LBP). The resulting features are used as an input for an SVM, which led to an accuracy of 93% for a dataset of 340 images (half of them ulcers).

In 2018, CNNs were used to automatically detect ulcers in WCE images for the first time [115]. In this work, an AlexNet architecture was used and achieved an accuracy of 95% in a dataset containing 3250 images with ulcers and 5000 normal images. Another approach using DL was presented in [116], where both GoogleNet and AlexNet were compared for the task of ulcer detection. Both the architectures achieved an accuracy of 100% in a dataset of 525 images, but is important to have into consideration that the test set had only 105 images (80 abnormal and 25 normal images), which is rather low for DL classifiers.

3.3.6. Multi-pathology Detection

There has been some work along the last years where a group of 2 different lesions has been detected with the same or similar methods. For example, in [104] both ulcers and polyps are detected, but the detection is done separately from one another. Besides the applied methods having small variations, two separate datasets were used (each with their type of lesion). In [111], both bleeding and ulcers are detected with the same method, but as in the previous one, the lesions are never tested jointly, being difficult to understand if the method can differentiate bleeding from ulcer.

Joining different types of lesions in the same dataset and classifying them all together has been a focus of a limited number of papers. For example, in [117] a CNN classifier was used in a

dataset containing 50 bleeding, 27 angioectasias, 8 chylous and 9 lymphangiectasias images. This method was used both for the task of segmentation and the task of image classification. In segmentation, a DICE of between 0.76 and 0.94 was achieved for the different lesions, while in classification only 1 image was wrongly classified. It is noticeable that the size of the dataset for classification is of a low size to correctly infer these results for clinical practice.

In [118], a CNN with a multi-scale pooling method is used for classification of different lesions (bleeding, erosions, colitis and gastricism), in a proportional dataset of 1560 images also with normal images. This method achieved an overall precision of 99%, going from 97% to 100% in the different types of lesions.

3.4. Conclusion

This chapter presented some of the methods applied in WCE exams to improve the physicians' analysis. The methods presented were divided into three groups: methods to reduce the size of the video; methods to improve the quality of the video and methods to automatically detect specific lesions. In the last group, different lesions were included in the analysis and to concluded, an overview of methods applied to scenarios where multiple pathologies were present.

Chapter 4. Methodology

Tis chapter includes an overview of the methods used in this work. Firstly, the analysis of color spaces in the different types lesions found in WCE exams and the best features to describe them. Secondly, the segmentation method used to improve the detection will be presented. Finally, the machine learning techniques used in this work will be explained, beginning with the single-learning models (MLP and SVM), passing by the Ensemble Learning technique, and finally a presentation of Deep Learning techniques used for the Multi-pathology challenge.

4.1. Characterization of Lesions and Features

Color is one of the most important characteristics in an image. It is by color we can distinguish different objects and different structures in a scene and specially in the GI tract, different lesions appear with different colors or different color features in WCE images, as seen in Section 2.1.1.

4.1.1. Color Spaces

The choice of the color spaces and channels used to analyze images is a difficult task, since it can directly influence posterior processing of these images. Due to its capability to enhance the image analysis process, color image segmentation is most widely used than the gray scale image segmentation [119]. A color space is a method by which is possible to specify, create and visualize color; RGB, HSV and CIELab are three of the most frequent chosen color spaces.

RGB is the standard color space, since it is the model used to acquire images from camera systems. This color space has three components: Red, Green and Blue, representing the color in each pixel by the amount of each one of the primary colors. Although being the most standard color spaces, it doesn't take into account the human vision perception regarding color distinction [120]. Also, the three components that together create this color space are highly correlated, making it difficult for individually process each component separately [64].

In Figure 9 the RGB model is represented into a cube, where each of the components (red, green and blue) is mapped in each of the vertices.

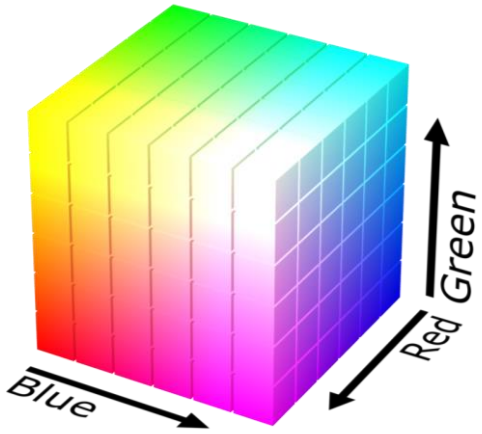


Figure 9. RGB color space mapped into a cube, where each vertex represents each component. Adapted from [121].

To overcome RGB color space non-linearity, color spaces which are based on human visual perception were created. One of them is HSV (Hue, Saturation, Value), which is a widely used color space in image processing when the aim is to automate a method the same way as the human eye performs it. This color space is usually represented by a cone, as shown in Figure 10. In this color space, H represents an angle relative to the red axis (the red color is considered to have and angle 0 and again at 2π). S component, describes how pure the hue is with respect to a white

reference. The value can be considered as the purity of color and is measured as a radial distance from the central axis (*0*) to the outer surface (maximum). The V channel represents the amount of lighting that illuminates a color [122]. Apart from its potential and as V channel takes into account the lighting levels of the region, it was proved that there is not enough independence between the channels of this space to be able to consider only the H and S channels without losing color information [123]. Its advantage lies in an extremely intuitive manner of specifying color [124]. An RGB image can be transformed into an HSV image using the following expressions, where the values of RGB channels lies between 0 and 1:

$$\mathrm{V} = \max(\mathrm{R}, \mathrm{G}, \mathrm{B}) \tag{1}$$

$$\mathrm{S} = \begin{cases} \mathrm{V} - \min(\mathrm{R}, \mathrm{G}, \mathrm{B}) & , \text{if } \mathrm{V} \neq 0 \\ 0 & , \text{otherwise} \end{cases} \tag{2}$$

$$\mathrm{H} = \begin{cases} 60\dfrac{\mathrm{G} - \mathrm{B}}{\mathrm{S}} & , \text{if } \mathrm{V} = \mathrm{R} \\ 120 + 60\dfrac{\mathrm{B} - \mathrm{R}}{\mathrm{S}} & , \text{if } \mathrm{V} = \mathrm{G} \\ 240 + 60\dfrac{\mathrm{R} - \mathrm{G}}{\mathrm{S}} & , \text{if } \mathrm{V} = \mathrm{B} \end{cases} \tag{3}$$

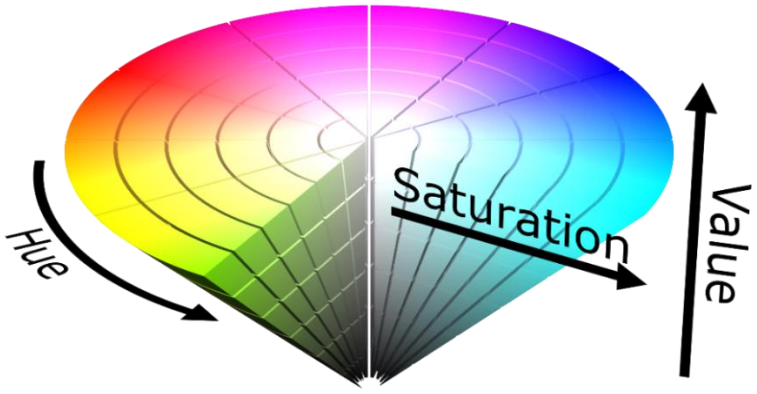


Figure 10. HSV color space represented as a conical object. Adapted from [121].

While HSV is very useful to separate hue and saturation of a color, the value component (the amount of light) is a relative brightness, which can collide with specific applications where this physical attribute is useful to be considered [125]. Contrarily, CIELab is a three-dimensional uniform model (Figure 11), which is represented by three different channels. The lightness information (*L*) goes from 0 (black) to 100 (diffuse white) and represents the absolute brightness. The other two channels are the color-opponent components *a* and *b*, which don't have theoretical

limits. Negative values of *a* channel indicate green and positive indicate red/magenta; while negative values of *b* channel indicate blue and positive indicate yellow. *Lab* color space was designed to approximate human vision, once its *L* component closely matches human perception of lightness [6], [126]. This color space can be used to make accurate color balance correction by modifying *a* and *b* components, or to adjust the lightness contrast using the *L* component [127]. Since this channel is independent from the others is possible to manipulate it without losing any color information [119]. Images can be converted from RGB color space to CIELab in an indirect way; this conversion process is done in two steps: conversion from RGB to XYZ, followed by the conversion to the CIELab color space. This process can be performed by applying the following equations [128]:

$$\begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.755160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{4}$$

$$\begin{aligned} \mathrm{L} &= 116\mathrm{f_y} - 16 \\ \mathrm{a} &= 500\left(\mathrm{f_x} - \mathrm{f_y}\right) \\ \mathrm{b} &= 200\left(\mathrm{f_y} - \mathrm{f_z}\right) \end{aligned} \tag{5}$$

$$f_x = \begin{cases} \sqrt[3]{\frac{X}{X_r}}, & \frac{X}{X_r} > \varepsilon \\ \frac{k\left(\frac{X}{X_r}\right) + 16}{116}, & \frac{X}{X_r} \leq \varepsilon \end{cases} \tag{6}$$

$$f_y = \begin{cases} \sqrt[3]{\frac{Y}{Y_r}}, & \frac{Y}{Y_r} > \varepsilon \\ \frac{k\left(\frac{Y}{Y_r}\right) + 16}{116}, & \frac{Y}{Y_r} \leq \varepsilon \end{cases} \tag{7}$$

$$f_z = \begin{cases} \sqrt[3]{\frac{Z}{Z_r}}, & \frac{Z}{Z_r} > \varepsilon \\ \frac{k\left(\frac{Z}{Z_r}\right) + 16}{116}, & \frac{Z}{Z_r} \leq \varepsilon \end{cases} \tag{8}$$

where X_r, Y_r and Z_r are the correspondent tristimulus values of X, Y and Z of a perfect white sample under the chosen illuminant, while ε and k are constants defined by the CIE (International Commission on Illumination) standards.

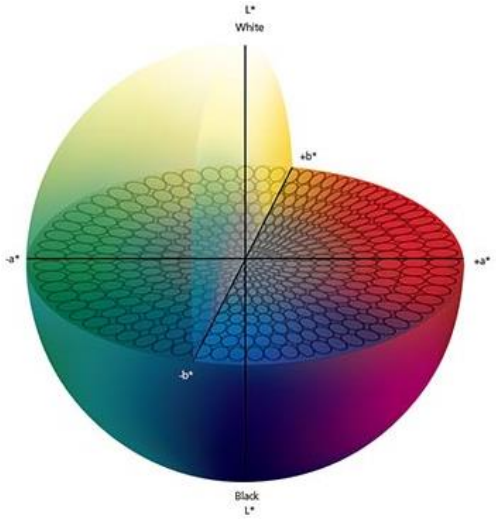


Figure 11. CIELab color space represented in a sphere. Adapted from [129].

4.1.2. Tissue Characterization

The different color spaces and their components highlight different structures in the tissue, therefore each type of lesion appears differently in the different color channels. Figure 12-16 (a) show different lesions (ulcer, angioectasia, bleeding, polyp and tumor, respectively) in the different components: (b) R, (c) G and (d) B channels from RGB; (e) H, (f) S and (g) V channels from HSV and (h) L, (i) a and (j) b channels from CIELAB.

It is clear that, while some of the lesions appear highlighted in some of the channels, in others this does not happen. For example angioectasias (Figure 13) or bleeding (Figure 14), which are blood-related lesions and consequently have a reddish appearance, appear highlighted in the channel *a* of CIELab (sub-figure (i)); which is expectable since high values of this component represent the color red (as seen in Figure 11). Ulcers (Figure 12) are white and have a red coloration around the lesion, and looking at the different channels, it is clear that the lesion itself appear highlighted in several of them. Although, only in channel a both the lesion and the inflamed area around it can be easily visually distinguished. When looking at polyps (Figure 15), these lesions are not clearly distinguishable with color differences; which can be seen in the various channels. In this case, the features that can differentiate them would be texture features. In the case of tumors (Figure 16), although texture features are also an important characteristic, there are also color characteristics important to characterize them. The three components from HSV and *a* and *b* channels from CIELAB seem to be appropriate to differentiate tumoral tissue.

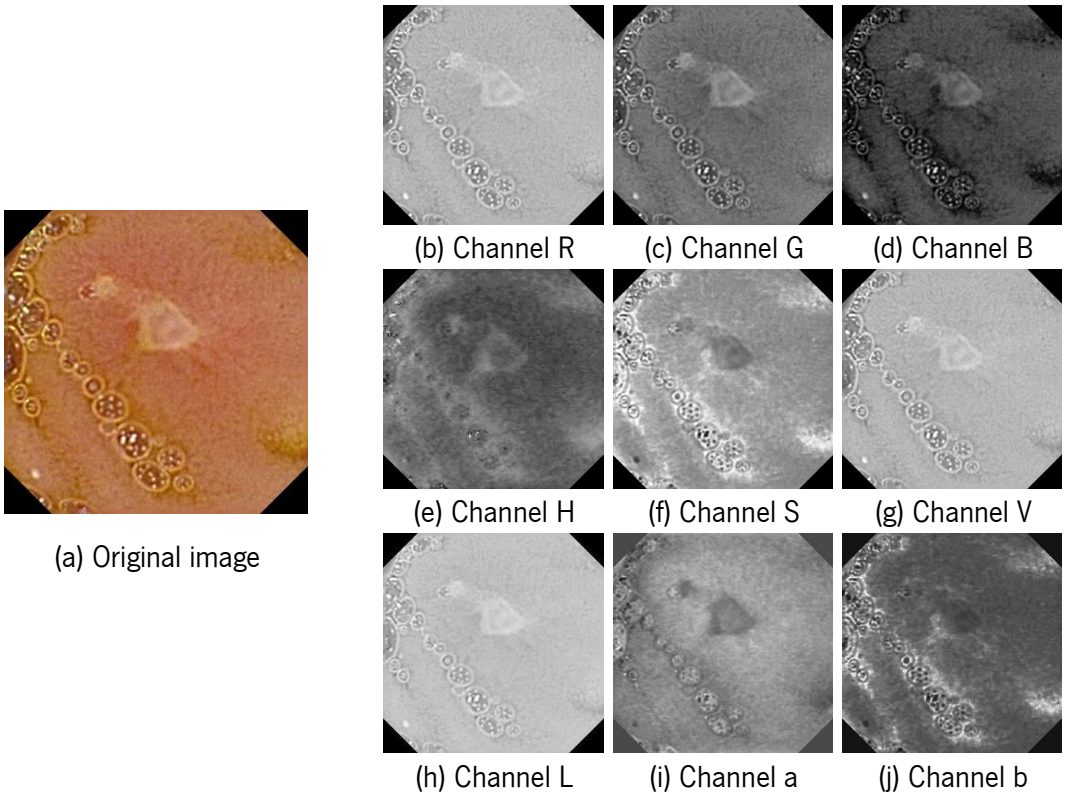


Figure 12. (a) Image with an ulcer, (b)-(d) Channels from RGB, (e)-(g) Channels from HSV and (h)-(j) Channels from CIELab.

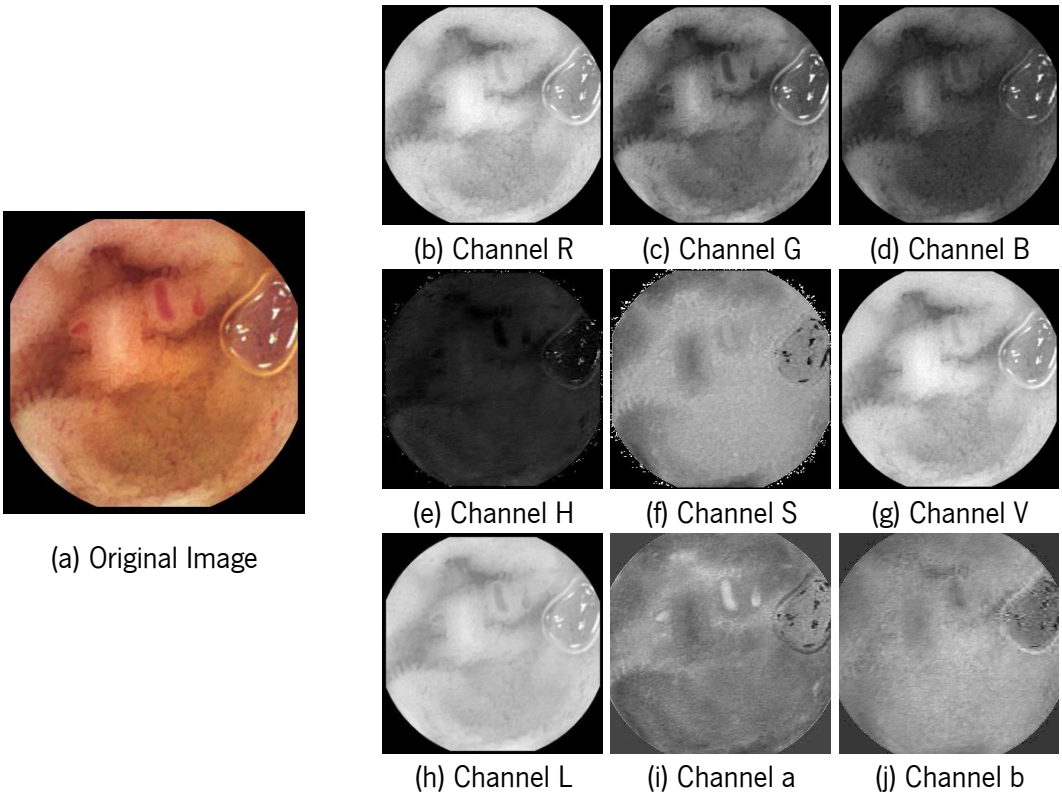


Figure 13. (a) Image with an angioectasia, (b)-(d) Channels from RGB, (e)-(g) Channels from HSV and (h)-(j) Channels from CIELab.

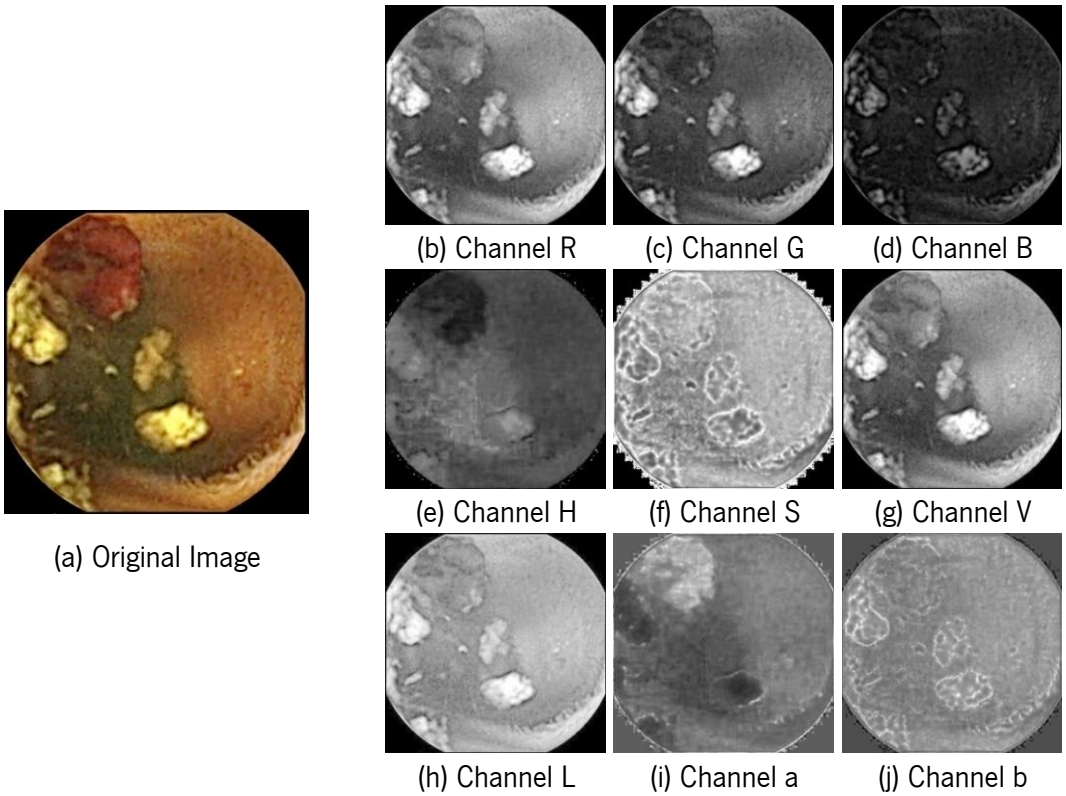


Figure 14. (a) Image with a bleeding, (b)-(d) Channels from RGB, (e)-(g) Channels from HSV and (h)-(j) Channels from CIELab.

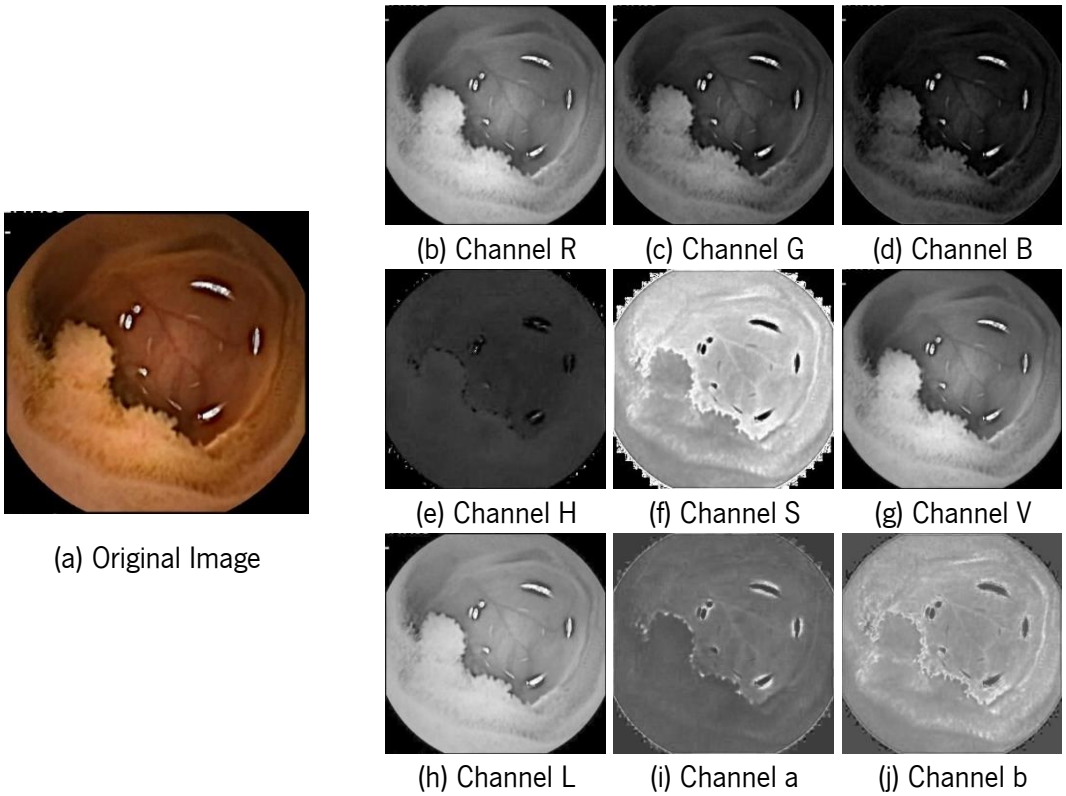


Figure 15. (a) Image with a polyp, (b)-(d) Channels from RGB, (e)-(g) Channels from HSV and (h)-(j) Channels from CIELab.

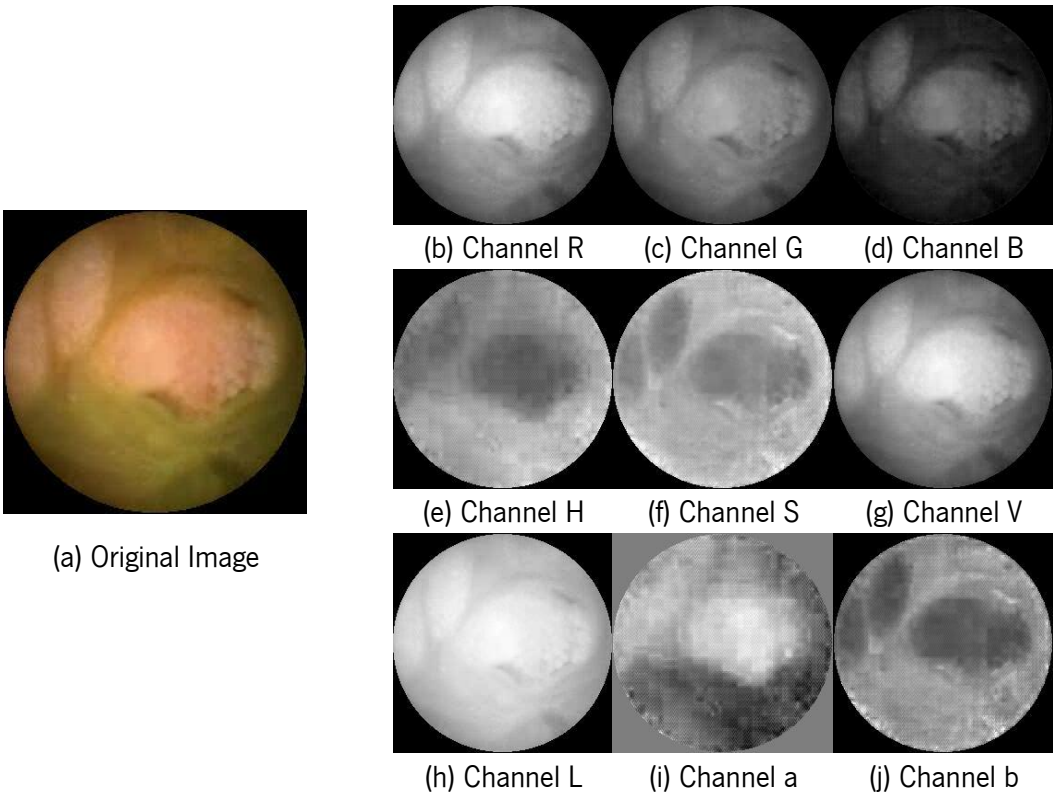


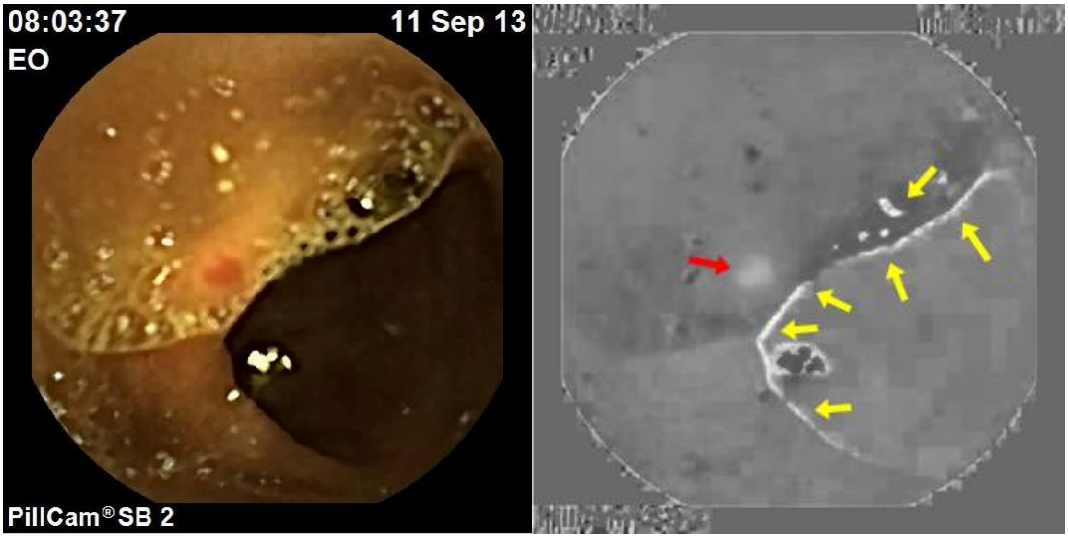
Figure 16. (a) Image with a tumor, (b)-(d) Channels from RGB, (e)-(g) Channels from HSV and (h)-(j) Channels from CIELab.

It is clear looking by at the previous images that the RGB color space is not appropriate to define and characterize GI lesions in WCE images. Also, textural based features can add significant information to the color-based ones, hence they should be used for all types of lesions.

4.1.3. Pre-Processing

One of the most important issues in the processing of WCE exams is the frequently appearance of non-expected and well-defined structures such as digestive fluid in the form of bubbles. These structures change color and texture of the underlying tissue, so they can affect the recognition system in structures not relevant for tissue characterization. Figure 17 shows a frame with an angioectasia lesion (a) and the correspondent component *a* of the same frame (b). As was seen in the previous section this is the channel where this type of lesions appears most highlighted. Looking at sub-figure (b) it is clear that the angioectasia tissue (red arrow) is highlighted when comparing with normal tissue. Nevertheless, the set of pixels next to the yellow arrows also appear highlighted, not corresponding to angioectasia tissue. These regions correspond to shadows of

some bubbles; which are characterized by low levels of intensity for green and/or blue components which usually don't characterize tissue regions.



(a) (b)

Figure 17. Image showing non-tissue regions. (a) Original image and (b) channel *a* from CIELab.

Usually, in regions where small bowel tissue is present, intensity values of RGB color channels do not approach from zero. This happens only in obscure regions or other non-informative and non-lesion structures. In this way, a pre-processing algorithm was developed to improve the quality of *a* channel to distinguish lesions (Algorithm 1). In here, C is a RGB image with a size $M \times N$ and D the same image, but in the CIELab color space. $C^k(i,j)$ and $D^l(i,j)$ represent the corresponding pixel in the component $\mathrm{k} = \mathrm{R}, \mathrm{G}, \mathrm{B}$ and $\mathrm{l} = \mathrm{L}, \mathrm{a}, \mathrm{b}$ and with the coordinates $\mathrm{i} = 1, 2, \ldots, M$ and $\mathrm{j} = 1,2, \ldots, \mathrm{N}$. In this algorithm the pixels that present values of green or blue components lower than a chosen threshold (δ), are replaced by an average of a neighboring region (with a variable size) centered in that pixel ($\mathcal{N}\{D^{\mathrm{l}}(\mathrm{i},\mathrm{j})\}$).

Algorithm 1: Pre-processing Algorithm for bubbles extraction

For each pixel (i,j)

  If $C^G(i,j) <$ or $C^B(i,j) <$

    Set $D^a(i,j) \leftarrow \mathcal{N}\{\mathrm{D^a(i,j)}\}$

  End If

End For

This algorithm was applied to images with angioectasia (since these lesions appear with high values of channel *a* from CIELab) and results are shown in Section 5.2.2.

4.2. Extraction of Features

One of the most important aspects in machine learning is feature extraction. An image containing GI tissue, and specially a lesion, can be characterized by several descriptors, such as color, texture, shape and size [130]. Although features of these four types can be used for tissue characterization, in the context of WCE lesion detection only color and texture will be used in this work, since both shape and size vary a lot.

4.2.1. Histogram-based features

Among visual features, color is one of the most vital, reliable and widely used. During this work, a probabilistic method was used to perform segmentation in the images (as described in Section 4.3.), so it is reasonable to use statistical measures from the histogram of the different channels to characterize a region. The shape of the histogram provides many clues to the characteristics of the image. For example, a narrowly distributed histogram indicates a low-contrast image. A bimodal histogram often suggests that the image contains an object with a narrow intensity range against a background of differing intensity. Different useful parameters can be worked out from the histogram to quantitatively describe a data distribution, such as mean (μ_t), which describes the centrality of the distribution, variance (σ_t) and entropy (H_t), which describes the dispersion of the distribution and also kurtosis (k_t), which represents the values of the distribution's tails [131], [132]. Some of these measures can be obtained as follows:

$$\mu_t = E\{X_t\} = \frac{1}{N_t}\sum_t x_t \tag{9}$$

$$\sigma_t^2 = E\{(X_t - \mu_t)^2\} = \frac{1}{N_t}\sum_t (x_t - \mu_t)^2 \tag{10}$$

$$H_t = E\{-\ln p(X_t)\} = \sum_t p_t \ln p_t \tag{11}$$

$$k_t = \frac{E\{(X_t - \mu_t)^4\}}{\sigma_t^4} = \frac{\frac{1}{N}\sum_t (x_t - \mu_t)^4}{\left(\frac{1}{N}\sum_t (x_t - \mu_t)^2\right)^2} \quad (12)$$

where X represents a random variable and x_t a particular value of X. N_t is the total number of samples in the vector X.

When using different features extracted from different color channels a features set is built. One alternative to statistical characterize each segmented region, which is usual in state-of-the-art systems, may be the cross characterization of features from the different regions. This cross characterization can be done by covariance measures if only second order statistical measures are used. Covariance among statistical measures can be computed as shown in equation (13).

$$\phi(i)_{R_x,R_y} = E\left\{\left(X_i^{R_x} - \mu_{X_i^{R_x}}\right).\left(X_i^{R_y} - \mu_{X_i^{R_y}}\right)\right\} \quad (13)$$

where X is a feature vector with the values for all the color channels considered, R are the different regions of the image (random variables), and i each statistical measure.

One of the main advantages of this approach is that relative measures regarding both regions are more effective than absolute measures that appear corrupted by random processes such as lightening differences among different devices and color differences among different subjects. Consequently, it is expected that differences between both regions are more device and subject independent than its absolute measures. Experimental results showed the effectiveness of this approach.

4.3. Segmentation

There are a lot of segmentation algorithms that can be used to separate an image into a number of smooth intensity regions. In this work, the segmentation module is based on the MaP approach by using the EM algorithm. A modified version of the Andersen acceleration algorithm is proposed in order to make the EM usable in poor separated cases. A modified MRF, with a weighted boundary function, was included for spatial context modeling purposes.

4.3.1. Expectation-Maximization

Regarding region-based image segmentation the most used statistical model is the Gaussian Mixture Model (GMM). The number of mixture components is usually the number of expected classes. In this case, as the purpose is to divide the abnormal from the normal tissue, the number of classes is set to 2. Because in this case each observation is the intensity of each pixel in each color channel, if more than one of these components is considered in the problem, the GMM is in fact a multivariate GMM [133].

The segmentation module uses a statistical classification based on Bayes rule (Eq. (14)). This rule indicates how the posterior probability of each class is calculated [134].

$$\mathrm{p}(\omega_i|x_j) = \frac{p(x_j|\omega_i).p(\omega_i)}{p(x_j)} \tag{14}$$

In this equation, x_j refers to the jth class and x to the feature vector, while $p(\omega_i|x_j)$, $p(x_j|\omega_i)$ and $p(\omega_i)$ are the a posteriori probability of class ω_i, the class conditional probability density function and the a priori probability, respectively. The term $p(x_j)$ is a scaling factor with the purpose of setting the *a posteriori* probability to the range between zero and one (and keeping $p(\omega_i|x_j)$ as a true probability value). This scaling factor can be ignored since it takes the same value for each class x_j. Therefore, for comparison purposes among classes, Eq. (14) can be written as:

$$\mathrm{p}(\omega_\mathrm{i}|x_j) = \mathrm{p}(x_j|\omega_i).\mathrm{p}(\omega_i) \tag{15}$$

Eq. (15) is used in the MaP estimate, which can be considered, assuming only two classes, as the following:

If $\mathrm{p}(x_j|\omega_1).\mathrm{p}(\omega_1) > \mathrm{p}(x_j|\omega_2).\mathrm{p}(\omega_2)$, then x_j belongs to ω_1,

otherwise x_j belongs to ω_2.

Since these models are not known for small bowel tissues they are required to be estimated. This estimation procedure is usually achieved by using the EM Algorithm.

To estimate $p(x_j|\omega_i)$ the well-known EM algorithm is used [135]. The main idea of this algorithm is to iteratively find the most appropriate parameters of the GMM model according to the

maximum likelihood (ML) criterion. The likelihood of an observation vector regarding the GMM is given by:

$$\mathrm{p}(\mathrm{x_j}|\varphi_\mathrm{i}) = \sum_{\mathrm{i}=1}^{\mathrm{k}} p(\omega_\mathrm{i}) \,.\, \mathrm{p}(\mathrm{x_j}|\omega_\mathrm{i}, \varphi_\mathrm{i}) \tag{16}$$

where φ_i is the vector containing the parameters of the distribution of class ω_i [135] and k is the total number of considered classes. The *a priori* probability ($\mathrm{p}(\omega_\mathrm{i})$) has a precise meaning in the model regarding data partition over all classes, however it is frequently used as a spatial regularizer by capturing neighboring information, not taken into consideration in the Gaussian mixture model that models pixel intensities as random variables independent and identically distributed. The likelihood of the whole image (X), which represents the set of all the samples ($x_0, x_1, \ldots, x_n$), can be computed as:

$$\mathrm{p}(X|\varphi) = \prod_{j=1}^{n} p(x_j|\varphi) \tag{17}$$

Maximizing the likelihood can be achieved indirectly by maximizing the log-likelihood, since the logarithm is a crescent function. Apart from that, having Gaussian functions involved using the logarithm is advantageous regarding derivative calculations since the exponential is annulated. The log-likelihood of the whole image is given by:

$$\begin{aligned} \mathrm{l}(\varphi) = log p(X|\varphi) = log \prod_{j=1}^{n} p(x_j|\varphi) = \sum_{j=1}^{n} log p(x_j|\varphi) \\ = \sum_{j=1}^{n} log \sum_{i=1}^{k} p(\omega_i) \,.\, \mathrm{p}(x_j|\omega_i, \varphi_i) \end{aligned} \tag{18}$$

Maximization of this log-likelihood function requires the application of the gradient operator [135], which results in:

$$\begin{aligned} \nabla_{\varphi_k} \mathrm{l}(\varphi) = \sum_{j=1}^{n} \frac{1}{p(x_j|\varphi)} \,.\, \nabla_{\varphi_k} \left[\sum_{i=1}^{k} p(\omega_i) \,.\, p(x_j|\omega_i, \varphi_i) \right] \\ = \sum_{j=1}^{n} \frac{1}{p(x_j|\varphi)} \nabla_{\varphi_k} [p(\omega_k) . p(x_j|\omega_k, \varphi_k)] \end{aligned} \tag{19}$$

The log-likelihood function can be written only in terms of *a priori* and *a posteriori* probabilities and probability density function of each class and data vector:

$$\begin{aligned}\nabla_{\varphi_k} l(\varphi) &= \sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k).\frac{\nabla_{\varphi_k}[p(\omega_k).p(x_j|\omega_k,\varphi_k)]}{p(x_j|\omega_k,\varphi_k).p(\omega_k)} \\ &= \sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k).\nabla_{\varphi_k}\log[p(x_j|\omega_k,\varphi_k).p(\omega_k)]\end{aligned} \tag{20}$$

Maximizing the likelihood of the data requires zeroing the gradient. Knowing that the probability density function (pdf) of a multivariate normal distribution is given by:

$$\mathrm{p}(x_j|\omega_k,\varphi_k) = \frac{1}{(2\pi_k)^{\frac{D}{2}}.|\Sigma_k|^{\frac{1}{2}}}.\exp\left(-\frac{1}{2}(x_j-\mu_k)^T.\Sigma_k^{-1}.(x_k-\mu_k)\right) \tag{21}$$

where D represents the dimension of the distribution. If $D = 1$, which is the case where only one color component is considered, the pdf function can be simplified as the following:

$$\mathrm{p}(\mathrm{x_j}|\omega_\mathrm{k},\varphi_\mathrm{k}) = \frac{1}{\sqrt{2\pi_k\sigma_k}}.\exp\left(-\frac{(x_j-\mu_k)^2}{2\sigma_k^2}\right) \tag{22}$$

The update of the distribution parameters (mean, covariance matrix and *a priori* probability) are given by:

$$\hat{\pi}_\mathrm{k} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{p}(\omega_\mathrm{k}|\,\mathrm{x_j},\varphi_\mathrm{k}) \tag{23}$$

$$\hat{\mu}_\mathrm{k} = \frac{\sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k).x_j}{\sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k)} \tag{24}$$

$$\hat{\Sigma}_k = \frac{\sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k)(x_j-\mu_k)(x_j-\mu_k)^T}{\sum_{j=1}^{n} p(\omega_k|x_j,\varphi_k)} \tag{25}$$

The weight of class j shown in Eq. (23) can be affected by pixel neighborhood, using MRFs, which will be explained in a later subsection.

Algorithmically we have an iterative process that maximizes the likelihood of the data. Initial estimates are however required and can be given by the K-means algorithm, which divides the pixels into k clusters in such a way that each pixel will be in the cluster with the nearest mean of the pixel value. Sometimes the convergence of this solution can be slow and hard to get. This

happens because some data is missing (the class to which each pixel belongs is unknown), so no closed form solution exists.

The EM has two main steps:

E-step (expectation): according to previous estimated parameters, the likelihood of each sample for each cluster is computed.

M-step (maximization): each pixel is associated with the cluster for the which the likelihood is higher. When this is done for all samples, new estimates of the model parameters are calculated.

After the M-step, the algorithm is repeated until convergence is achieved. Convergence is achieved when the content of each cluster doesn't change in consecutive iterations [136].

4.3.2. Anderson Algorithm

One of the difficulties encountered in the re-estimation procedure inherent to the EM algorithm is that sometimes the convergence is slow, especially when classes are poorly separated. This can have negative implications in the use of the algorithm for cases where thousands of frames need to be segmented. Hence acceleration procedures are of utmost importance in the current context. Anderson acceleration algorithm is based on the concept of fixed-point iteration whose goal is to solve iteratively $g(x)=x$. The general algorithm can be described by:

Algorithm 2: General Fixed-Point Iteration

(1) Choose an initial guess $x_0 \in R^d$.

(2) Specify an error tolerance $\varepsilon > 0$ and a maximum iteration number I.

(3) Iterate.

For $\mathrm{k} = 1,2, \ldots, \mathrm{I}$

Set $x_k = \mathrm{g}(x_{k-1})$

If $\|x_k - x_{k-1}\| \leq \varepsilon$

Return x_k

Break

End If

End For

The algorithm can be used for maximization procedures through the relationship f(x)=g(x)-x=0. It is shown [137] that the MLE is a fixed-point of the EM estimation map. The basic idea of the Anderson acceleration algorithm is to make use of information gained from previous iterations to predict the next one. Therefore in its more general form the algorithm as given in [138] is as follows:

Algorithm 3: Anderson Acceleration ALgorithm

(1) Given x_0, $m \geq 1$ and a maximum iteration number I.

(2) Set $x_1 = g(x_0)$.

(3) Iterate.

For $\mathrm{k} = 1,2, \ldots, \mathrm{I}$

Set $m_k = \min \mathrm{mk}$

Set $\boldsymbol{F_k} = \left(\boldsymbol{f_{k-m_k}}, \ldots, \boldsymbol{f_k}\right)$ where $f_j = \mathrm{g}\left(x_j\right) - x_j$.

Find $\boldsymbol{\alpha^{(k)}} = \left(\boldsymbol{\alpha_0}^{(\boldsymbol{k})}, \ldots, \boldsymbol{\alpha_{m_k}}^{(\boldsymbol{k})}\right)^{\boldsymbol{T}}$ that solves

$$\min_{\boldsymbol{\alpha}=\left(\boldsymbol{\alpha_0},\ldots,\boldsymbol{\alpha_{m_k}}\right)^{\boldsymbol{T}}} \|\boldsymbol{F_k}\boldsymbol{\alpha}\|_{\boldsymbol{2}} \quad \text{subject to } \sum_{\boldsymbol{j=0}}^{\boldsymbol{m_k}} \boldsymbol{\alpha_j} = \boldsymbol{1} \text{ (*)}.$$

Set $x_{k+1} = \sum_{j=0}^{m_k} \alpha_j^{(k)} g\left(x_{k-m_k+j}\right)$.

End For

This constrained optimization problem (due to (*)) can be reformulated in an unconstrained optimization problem which solution can be found in [138]. However, the convergence of the EM algorithm is usually characterized by a smooth curve which means that last iterates can predict more accurately the next one than older iterates. Therefore, our implementation of the Anderson accelerated algorithm has two additional constraints, characterized by $\alpha_j^{(k)} < \alpha_{j+1}^{(k)}$ and $\alpha_{\mathrm{j}}^{(\mathrm{k})} > \delta$, where δ is a small constant that assures that all the previously selected iterations are effectively used. This constrained optimization problem can be solved by using state-of-the-art linear solvers as follow.

First a vector containing the differences of the log-likelihood between successive iterations must be composed. These differences capture the rate of convergence of the EM algorithm from iteration to iteration. Defining this vector as $\overline{c} = \{c_1, c_2, \ldots, c_n\}$ and assuming the existence of

the α vector such as $\overline{\alpha} = \{\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n\}$ the main goal is to find the alpha vector that minimizes the function:

$$f(\bar{\alpha}) = \sum_{i=1}^{n} c_i \alpha_i \tag{26}$$

subject to the following constraints:

$$\begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^{n} \alpha_i = 1 \\ \alpha_i \leq \alpha_{i+1} \end{cases} \Leftrightarrow \begin{cases} -\alpha_i \leq 0 \\ \sum_{i=1}^{n} \alpha_i - 1 = 0 \\ \alpha_i - \alpha_{i+1} \leq 0 \end{cases} \tag{27}$$

where the right-side equation (27) shows a more suitable format for purposes of the used optimization solver. The computed alpha values are then used to compute GMM updates shown in equations (23), (24) and (25), forcing the next iteration to contain convergence dynamics observed in the n last iterates.

4.3.3. Markov Random Fields

MRF models have the ability of capturing neighborhood information to improve a priori probabilities $p(\omega)$. An image can be considered as a random field, or a collection of random variables ($\Omega = \Omega_1, \ldots, \Omega_N$) that are defined on the set S. A random field is considered an MRF only when the following conditions are fulfilled:

1. $p(\omega) > 0, \forall \omega \in$, which is the condition of positivity;
2. $p(\omega_j | \omega_{S-\{j\}}) = p\left(\omega_j \middle| \omega_{N_j}\right)$, which is the condition of Markovianity.

The first condition is straightforward fulfilled just because the values are probabilities, so by definition, greater than 0. The condition of Markovianity states that the probability of an observation x_j, given the other random variables in the field, is equal to the probability of the same observation, given a neighborhood around its location ($\mathcal{N}_j$), or in other words, given its neighborhood, a variable is independent on the rest of the variables. Fulfilling this condition is in fact modelling the neighborhood effect. Using Gibbs Random Field (GRF), the *a priori* class probability can be assigned such as:

$$\mathrm{p}(\omega) = \frac{1}{Z}\exp\left(\frac{-U(\omega)}{T}\right) \tag{28}$$

$$\mathrm{Z} = \sum_{\omega} e\,xp\left(-\frac{U(\omega)}{T}\right) \tag{29}$$

In this equation, the constant T represents the temperature and controls the level of peaking in the probability density, and the quantity Z is a normalizing constant which guarantees that $\mathrm{p}(\omega)$ is always between zero and one. $\mathrm{U}(\omega)$ is an energy function and is obtained by summing all functions $V_C(\omega)$ (clique potential) over all C possible cliques. A clique is defined as a grouping of pixels in a neighborhood system, such that the grouping includes pixels that are neighbors of another in the same system.

$$\mathrm{U}(\omega) = \sum_{C\in\mathcal{N}} V_C(\omega) \tag{30}$$

The Hammersley-Clifford theorem defines that if and only if a random field Ω on S is a MRF with respect to neighborhood system $\mathcal{N}$, then Ω is a GRF on S with respect to a neighborhood system $\mathcal{N}$. This fact allows to convert the conditional probability as a Markovianity condition of a MRF to the non-conditional probability of a Gibbs distribution of Eq. (28).

To compute the estimation of $\mathrm{p}(\omega)$, the energy function used was based on Reference [139]:

$$\mathrm{U}(\omega_j) = \sum_{k} \beta_k . l_{k,j} \tag{31}$$

In Eq. (28), k is the direction (in this case it can be horizontal or vertical) and $l_{k,j}$ is the Dirac impulse function in such a way that $\mathrm{U}(\omega_j)$ depends on the count of pixels in neighborhood that do not belong to class j.

Usually, in practice, models are considered as isotropic, so the amount of variables to estimate is strongly decreased, becoming in this case β_k a constant. However, pixels near the borders are sometimes wrongly classified in the Gaussian Mixture especially due to the partial volume effect. Therefore, using the β_k parameter to model intensity differences in neighborhood pixels in order to reinforce border conditions has been used in several works where several functions have been suggested. The main idea is to set β_k in such a way that a direct interference on border location is achieved. Heuristically we want to avoid class j under situations of high

variance that usually appear near borders, even if a large number of pixels belong to class j. Under relative smooth conditions the border can also be present and can be detected by pixel intensity variations which occurs at corners of small structures. Some tests were conducted in order to get the best β_k for pixels on and near the border. The function given by Eq. (32) was found and was proposed in the ambit of this work:

$$\beta_k = \frac{\sigma_k}{1 + exp\left(-\sigma_k \frac{\sum_i^n |I_i - I_c| \mathrm{dist}\,(I_i, I_c)}{n}\right)} \tag{32}$$

In Eq. (32), β_k is dependent on the difference of intensities ($|I_i - I_c|$) of the neighbor in the direction k, but also of the distance between the pixel in the center and the neighboring pixel ($\mathrm{dist}(\mathrm{I_i}, \mathrm{I_c})$). The term σ is the standard deviation of the neighboring used in this case, which is a 2D-neighboring system of 8 pixels that can be seen in Figure 18, where the darker pixel is the current observation.

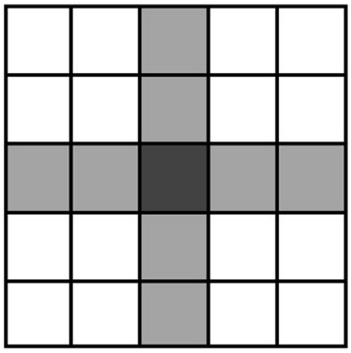


Figure 18. 2D-neighboring system of 8 pixels used in Eq. (32).

4.4. Machine Learning Methods

Pattern recognition is the application of methods with the purpose of finding patterns in data, which is of great importance in automatic diagnosis systems. These methods can achieve this result by teaching a certain model to detect these patterns, being called as Machine Learning methods. There are two main groups of machine learning systems: supervised and unsupervised. In supervised learning methods, a certain number of labeled examples of data are used for training purposes so the models can learn the best way to correctly separate the data into different classes (training phase). The algorithm will try to generalize its response to all possible inputs. After this,

usually there is a testing phase, where a set of examples never seen by the algorithm will be classified and its performance measure is based on the match of predicted and true labels. Unsupervised methods don't have a training phase, instead a function that characterizes data distribution is iteratively maximized until it best fits the available data (for example, the segmentation method explained in 4.3. is considered an unsupervised method). In this chapter, only supervised methods will be addressed.

4.4.1. Single Learning Models

The most important characteristics of supervised systems is the generalization capacity which represents the capability to produce sensible outputs for inputs that may not be part of the learning dataset. In the supervised algorithms, which can be regression or classification, there is a dataset (the training data) that consists of a set of input data that also has the target data (the answer that the algorithm should produce) [140]. Generalization capacity depends on training data, that can be chosen in such a way that all the training space is entirely represented and on training procedures especially at the validation level. In this work the conventional supervised classifiers used were the Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM).

4.4.1.1. Multilayer Perceptron

Neural Networks (NN) are computational models inspired by the behavior of biological neural systems. They consist in several processing elements (neurons) and connections between them (weights), a neural architecture, and a learning algorithm. A NN system offer a number of attractive properties and capabilities: non-linearity, input-output mapping, adaptivity, generalization, and robustness [141]. There are many types of NN architectures, but Feedforward NNs (FNNs), also known as MLPs, are the most popular and successful NN. They consist of one input layer, one or multiple hidden (intermediate) layers, and one output layer. In MLP, during prediction, the information propagates in only one direction (forward), from the input layer to the output layer. Unlike other statistical techniques, MLP makes no prior assumptions concerning the data distribution. It can model highly non-linear functions and can be trained to accurately generalize when presented with new, unseen data [142]. Single-hidden Layer Feedforward Networks (SLFNs) are MLP with just one hidden layer, and they are the simplest and mostly used FNN because of

the good approximation capabilities in many problems. The architecture of a single hidden layer MLP is shown in Figure 19.

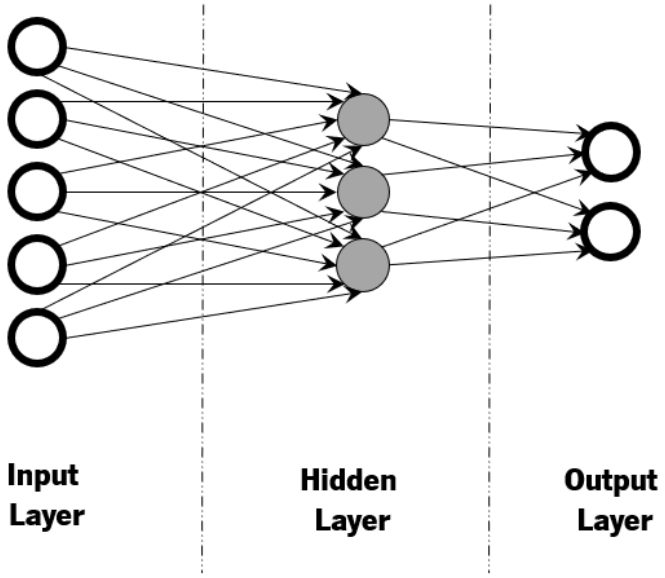


Figure 19. A representation of an MLP with 1 hidden layer.

In the example of Figure 19, each element of the feature vector is fed to each neuron of the input layer. After passing through the input layer, inputs are weighted and fed simultaneously to a second layer, until the final output layer is reached. The number of hidden layers is usually settled heuristically based on the ratio training examples/number of network parameters to be estimated (weights).

The weights of the neurons are adjusted in the training phase, using different learning algorithms. The most used is the backpropagation algorithm, which employs a gradient descent method to tune the weights of the MLP. The update of the network weights is an iterative process that guarantees a non-increasing value of the network error. The training is stopped when the output error drops below the predefined threshold or when a predetermined number of iterations is reached [141].

4.4.1.2. Support Vector Machine

Ever since its conception, the SVM classifier has been a prominent landmark in statistical learning theory and its success is attributed to two main premises: the original space is transformed into a very high-dimensional new space and a large margin can be found on this new space. Thus, it has the objective to define a hyperplane, which can be used for a good separation of the training

data. This good separation is achieved by the hyperplane that has the largest distance (margin) to the nearest training-data point of any class. The points which are at the exact minimum distance corresponding to the minimum margin are called the support vectors. These points are marked in Figure 20, and the dotted lines allows to visualize the margin of the hyperplane [140], [143]. The potential advantage of SVMs over MLPs is the control of the mentioned margin according to data variance which maximizes the range of data variation that preserves right classification. The drawback of SVMs is that the system tuning is usually hard to achieve since the handling of several parameters is usually required.

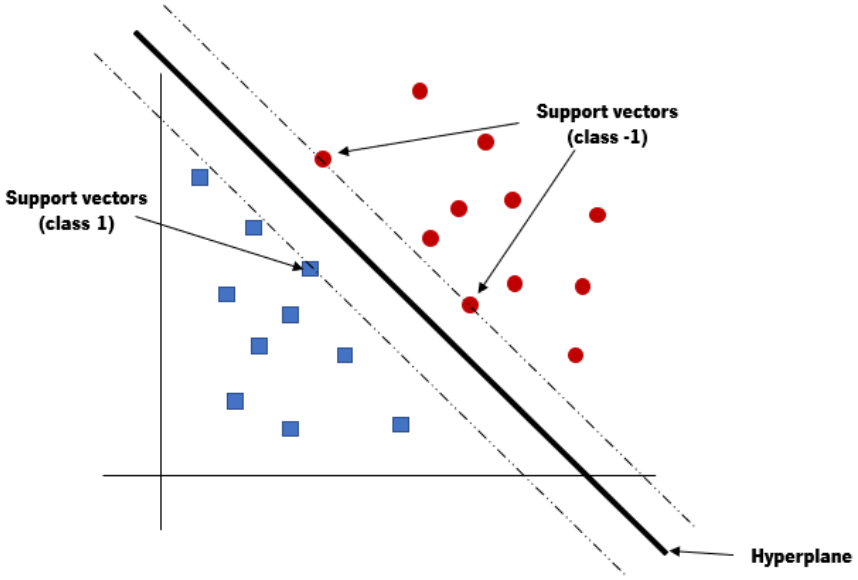


Figure 20. Representation of an SVM classification boundary (Hyperplane) and the support vectors, to divide two classes of data.

In Figure 20, a linear classification is performed, since the hyperplane used to divide the data is of linear nature. Although, SVM can also perform nolinear classification, which is quite useful in the majority of real life applications (where linear separable data is not common). This is done by mapping the non-linear separable data-set into a higher dimensional space where we can find a hyperplane that can separate the samples. This is performed using kernel functions, like polynomial, radial basis function, hyperbolic tangent, sigmoid, among others [140].

One of the main disadvantages of SVM classifiers is that the conventional classifier cannot deal with problems with more than two classes. In this specific case, a multiclass problem must be divided into multiple binary classification problems, and one SVM must be built for each one of the problems [143].

4.4.2. Ensemble Learning Models

Another way to perceive a classification task is to use several classifiers (of the same or different types) on the same classification task. EL systems are then sets of models that combine in their decisions, their learning algorithms, or different data to achieve accurate predictions [143]. These systems are usually more accurate than any single model used separately, and the effectiveness of ensemble systems has been shown in different benchmark data sets [141]. This classification strategy poses 2 main problems; which information must be aggregated and how this information can be aggregated.

Despite the good performance when using EL systems, this depends on several factors, such as the diversity between the models and the combination strategy of all the used models. Diversity of an EL system can be described as having differences between the different classifiers used in the ensemble. The major obstacle to achieve a high diversity lies in the fact that the individual learners are trained for the same task from the same training data, and thus they are usually highly correlated [143].

Some of the most used types of EL system configurations are bagging, boosting and stacking. Bagging applies bootstrap to randomly sample the training dataset, having in result a different training dataset for each classifier. Boosting also samples the training dataset, but trains the different models sequentially so the wrongly classified data can be improved in posterior classifications. When using stacking, the whole training dataset is used as input to all the classifiers; whose predictions are used in the final classifier reaching a final prediction [141], [143], [144].

4.4.2.1. Proposed Ensemble Classifier

The diversity of a classifier (which assures that each member of the ensemble is tuned for a subset of the training set) can be achieved by using bootstrapping, where the training data is randomly sampled with replacement. Then each bootstrapping subset is used for training each element of the ensemble. Bootstrapping does not assure coherence of the data, in the sense that similar data can be spread by different subsets. Consequently, data can also be spread by different elements of the ensemble, decreasing the diversity. Therefore, more thorough strategies are required in order to group similar data, such as the one used in [145], where the training set is partitioned into different subsets by using the Fuzzy C-means clustering algorithm. This approach

is employed in this work, where the training set is partitioned by using GMMs. Its parameters can be estimated by using the EM algorithm that in turn can be accelerated by using the modified version of the Anderson algorithm presented in Section 4.3.2. In fact, two GMMs are used: one for clustering normal data (M clusters) and another for clustering pathological data (N clusters). Different clusters in both classes will be combined to produce $M \times N$ training subsets. After partitioning the training data, each subset is employed to train a classifier which is a member of the ensemble. The classifiers used are Least Squares-SVM models.

Pathological and normal data are inherently of different nature, so each one need to be clustered independently generating a bi-dimensional structure of models. In other words, two GMMs are required for data partition purposes, one for each class; normal and abnormal. This reasoning can be extrapolated for the multi-pathology case where a N–dimensional structure of models is generated being N the number of classes.

Having generated a set of models with a high degree of diversity an appropriate combination strategy must be used. In fact, diversity is very important to take advantage of EL systems over single models however the accuracy can be seriously compromised since for a given input, the elements of the ensemble that never seen it before can generate inappropriate output values that can compromise seriously the global output of the ES. In this regard this combination strategy can be optimized according to the dataset by training procedures. This combination element is usually known in the current literature as gating network which purpose is to partition the input space into regions corresponding to the various ensemble elements. The gating network proposed is an MLP, perhaps currently the most used nonlinear classifier. A three-layer MLP with 3 neurons in the middle layer gave the best results for our current dataset, assuming a structure of 16 elements (4x4) and one output neuron. Figure 21 shows a modular view of the ensemble structure proposed for the case of two clusters for both pathological and normal data. The adopted terminology for model identification was X.Y where X and Y refers, respectively, to the cluster of pathological and normal data. So, for example, SVM1.2 refers to the model trained by using pathological data from cluster 1 and normal data from cluster 2.

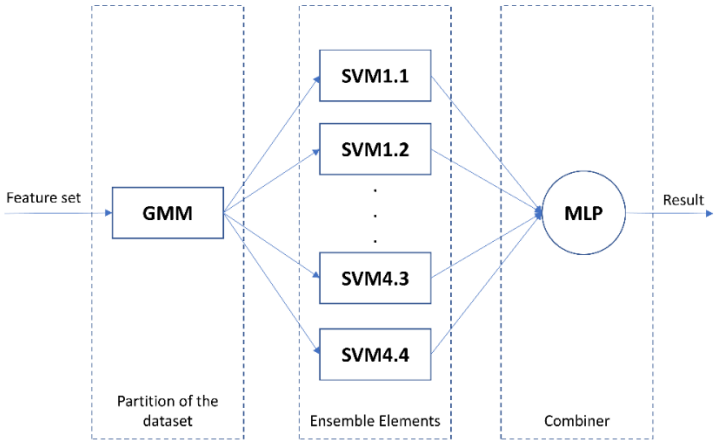


Figure 21. Block diagram of the proposed ensemble system for a feature set.

As the combiner is itself a classifier which goal is to assure high accuracy that is sometimes compromised by the high diversity achieved since no data is shared among the different ensemble elements, a modular train is required. Therefore, each SVM is trained by using each subset provided by the partition module (GMM). Then, the MLP (combiner) is trained by using the entire training set and keeping unchanged the SVMs so an optimum combination function is achieved. The architecture of the proposed ensemble is related with state-of-the-art mixtures of experts [146], although this includes a nonlinear module for subsets classification (SVMs) and one more powerful combining element. Additionally, as the WCE exam takes a long time and both tissue and light characteristics can change over time an environmental adaptation procedure can improve the system performance. This adaptation can be achieved by providing the system with current samples. In other words, the system is retrained with new samples that contain new environmental conditions. However, adapting ES decreases the diversity if the same data is shared by more than one element of the ensemble. Consequently, each new sample is pre-classified by the GMM module used for partitioning the training dataset. The Maximum Likelihood (ML) criterion is used to match each sample to the corresponding element of the ES, and only this element is adapted. This procedure assures system adaptation to new conditions, preserving the diversity.

4.4.3. Deep Learning Models

Deep Learning (DL) is considered as a family of machine learning classifiers which are based on learning several levels of representations, corresponding to a hierarchy of features, where higher-level concepts are defined from lower-level ones, and the same lower-level concepts can

help to define many higher-level concepts [147]. Most of these levels compute non-linear input-output mappings with the goal of transforming its input to be sensitive to most minute details and insensitive to irrelevant variations such as the background [148]. One of the most used networks in the DL methodology, are Convolutional Neural Networks (CNN), which are a type of deep networks for supervised learning.

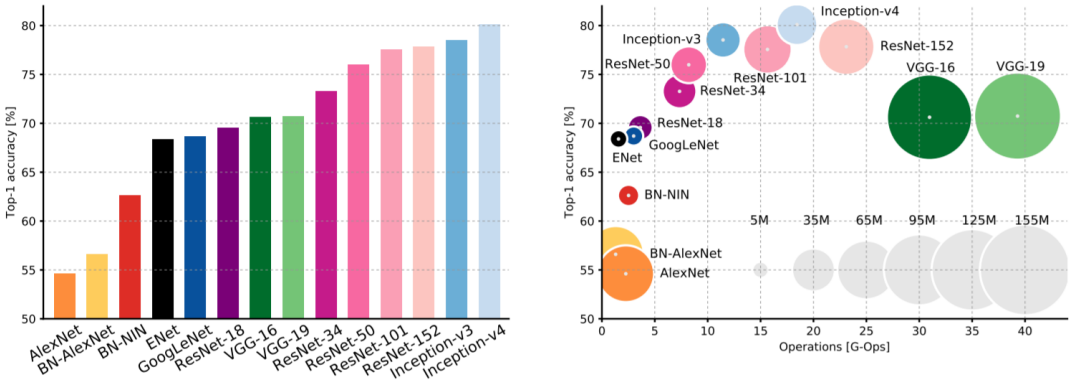
4.4.3.1. Convolutional Neural Networks

Each CNN consists in two different modules: a convolutional layer and a pooling layer. These modules are often stacked up with one on top of another, and while the convolutional layer shares many weights, the pooling layer subsamples the output of the convolutional layer and reduces the data rate from the layer below [147]. Deeper layers (which are closer to the input) will learn to detect simple features such as edges and color gradients, whereas higher layers will combine simple features into more complex features. Finally, dense layers at the top of the network will combine very high-level features and produce classification predictions. In a CNN, all low-level details are sent to the higher-level neurons. These neurons then perform further convolutions to check whether certain features are present. This is done by striding the receptive field and then replicating the knowledge across all the different neurons.

CNNs have some invariance properties (e.g., translation invariance), specially due to the weight sharing of the convolutional layers, and when appropriately pooling schemes are chosen. When using images as inputs, as is the case of this work, the result after passing through a first convolutional layer is an abstracted feature map with a shape that depends on the defined convolutional kernel. These models have been found highly effective and been commonly used in computer vision and image recognition for a wide-range of applications, from medical image [149], to traffic management [150] or handwriting recognition [151].

There are a high number of different architectures that can be used in CNNs, and each one has different properties and different performances when applied to classification tasks. Looking at the ImageNet classification challenge [152] (an annual challenge launched in 2010, where the purpose is to achieve the best performance in a multi-category classification task), it is possible to compare some of the most common CNN architectures. In Figure 22, two accuracy comparisons are shown, where several architectures were applied to the ImageNet classification challenge. This comparison was performed in 2017 and published in [153]. In Figure 22 (a), the

total accuracy is compared, showing that AlexNet has a worst accuracy while different versions of ResNet and Inception Networks show an improved accuracy. In Figure 22 (b) a different analysis is made, where not only the accuracy is considered, but also the computation cost and number of parameters of each network. One of the first conclusions that one can take is that VGG networks have the most expensive architectures (both in computation cost and number of parameters). It is also noticeable that AlexNet, although with the worst performance, has a high number of parameters (without increasing its computational cost). When looking at the rest of the networks, it is noticeable a proportion between the increase of the accuracy of the network, the higher number of parameters and more expensive computational requirements. Nonetheless, it is also noticeable that with the group of better performances (ResNet and Inception), it is not clear that an increase of complexity has a return in an increasing performance of the network, which can suggest that models somewhat reach an inflection point on this dataset [153].



(a) (b)

Figure 22. Comparison of different CNN's architectures. (a) Validation accuracies for top scoring single-model architectures. (b) Accuracy versus amount of operations required for a single forward pass, with the size of the blobs being proportional to the number of network parameters [153]**.**

4.4.3.2. Object Detection

The task of object detection is the one where any objects in an image frame are detected and correctly labeled, and usually occurs in two steps:

1. Object Localization, where a bounding box that encloses with the tightest possible option the region where the object is present.
2. Object Classification, where the objects found in the previous step are put into a classifier, that will give the right label to the object.

CNNs can be used for this purpose, specially the R-CNN and its successors, that will be briefly explained in the following sections.

R-CNN

The name R-CNN comes from the expression "Regions with CNN", and these networks have the purpose of detecting and localize objects in an image. This method intends to mimic the human visual cortex system, with a hierarchical and multistage process, to improve the classical CNN scheme to introduce the localization task in an image [154].

The localization of a specific region or object in an image can be seen as a regression problem, which turns out to be not effective for this task since it leads to poor results and a very slow process. Using a sliding-window detector could also be a good approach, but with the increase of subsequent layers, the resolution is lost, which leads to a lack of detail.

To overcome the limitations mentioned previously, R-CNN's were introduced by Girshick et al. [154], using an approach to detect and localize different structures without the need of slow detection speed. R-CNN have usually the following main modules in its structure (Figure 23):

1. **Region Proposals**: to generate a number of proposed bounding boxes;
2. **Feature extraction:** extraction of features from CNN that will help to classify each region
3. **Classification**: classification of the different regions using Linear SVM

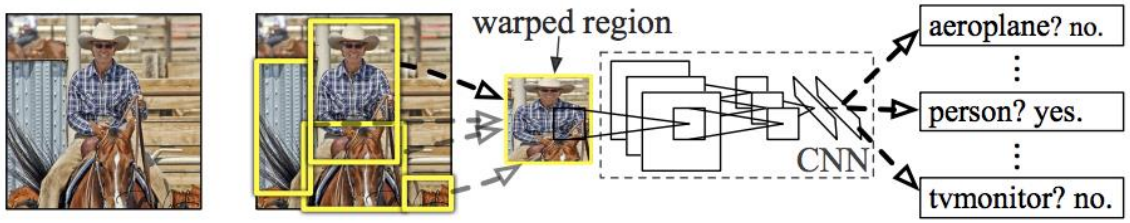


Figure 23. Scheme of R-CNN structure with its different modules. Adapted from [154]**.**

Looking at the different modules individually, first we have the Region Proposals module, which intends to select different RoIs (approximately 2000 candidates per image), which can contain different target objects and can be of different sizes. This selection is performed with a selective search method, which captures different scales within the image to find different regions with a diversity of measures [155]. In the Feature Extraction step, each region proposal is used as

input to the following CNN architecture, which originally used AlexNet. The output of this CNN are a group of features, that are used in a SVM in the third phase (Classification), which predicts if an object is present on each one of the region proposals. Although presented as a strong tool for object detection when it appeared, it still has some drawbacks, mainly regarding the long time to train the network, due to the high number of region proposals per image. Also, since the selective search method is fixed, there is no learning during that phase, and can lead to bad candidates in the pool of the region proposals.

Fast R-CNN

Due to the limitations of the R-CNN algorithm, the same authors proposed an improvement, which was named Fast R-CNN [156]. This method is similar to R-CNN in the overall methodology but fixes the problem around the speed of the original method. The architecture of Fast R-CNN is represented in Figure 24, where the similarities with Figure 23 are obvious.

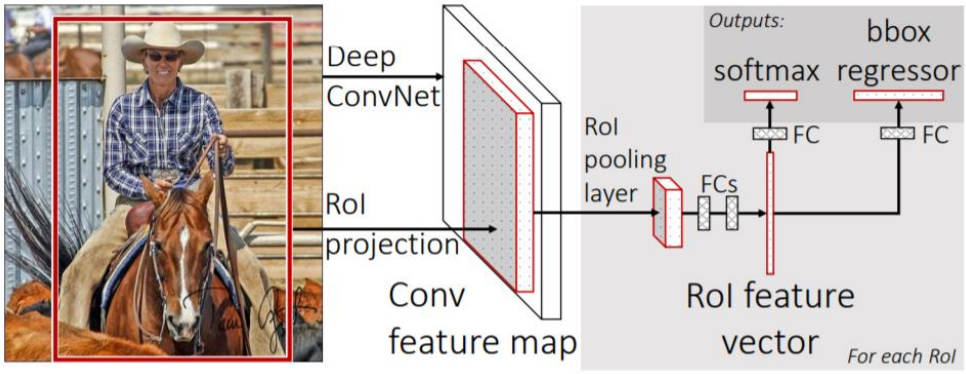


Figure 24. Architecture of the Fast R-CNN method. Taken from [156]**.**

In this new method [156], the original image is fed to the CNN to generate a convolutional feature map, which in turn is used to identify the region proposals. Afterwards, feature vectors selected for each region proposal are fed into the RoI pooling layer, which subsequently are put into a sequence of Fully Connected (fc) Layers. Afterwards, with softmax probability estimates, each region will be classified in terms of the object it contains, or not. An extra branch of the architecture, Bounding Box (BB) regression branch, is used to improve the limits of the region proposals, so they are more precise with the objects found.

There are two key differences between Fast R-CNN and its predecessor:

1. The 2000 region proposals are not fed to the CNN every time an image is used, since the whole image is used as input and only afterwards the different region proposals are fed to it.
2. The use of a softmax approach for object classification, instead of the standard SVM, which slightly improves the classification performance.

Indeed, authors shown improvements of more than 10 times regarding training time and more than 50 times regarding testing time. Regarding classification and segmentation performance, although slightly higher, they present similar values for mAP and accuracy [156].

Faster R-CNN

The Fast R-CNN method has still one main drawback, that is related to its region proposer, the selective search algorithm (which was inherited from its predecessor R-CNN). This method, although presenting good results, was rather slow, being considered as the bottleneck of the whole process. From this knowledge, it was proposed a new variant of this method, named Faster R-CNN, which uses a Region Proposal Network (RPN) instead of the Selective Search method [157]. The whole architecture of this system is shown in Figure 25.

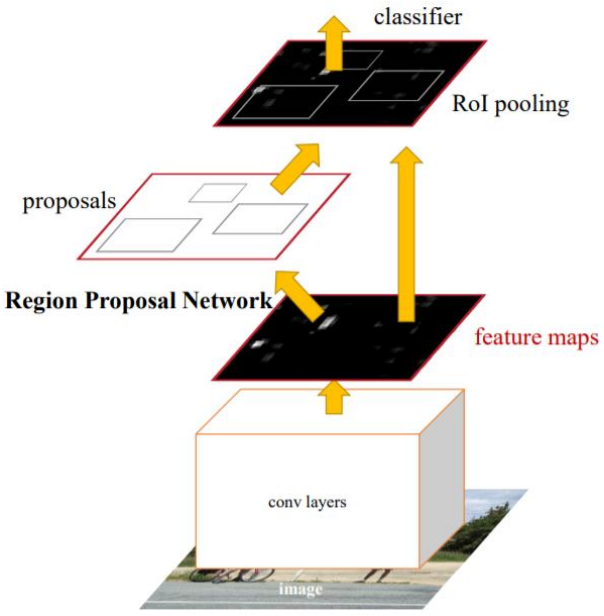


Figure 25. Architecture of the Faster R-CNN method. Taken from [157]**.**

Looking at this architecture, it is possible to understand that the same convolutional feature maps can be used to generate the region proposals. This is achieved by adding a Fully

Convolutional Network (FCN) on top of CNN features, creating this way a RPN. This network passes a slide window over the CNN feature map, retrieving a number of potential bounding boxes (together with a value representing the quality of each bounding box) within the image. These bounding boxes are then fed to the classifier used in the Fast R-CNN methodology, and consequently using its main advantages.

4.4.3.3. Instance Segmentation

Contrarily to the object detection task that was addressed in the previous section, the instance segmentation task intends to associate and classify each pixel to a determined class, and not just a region around a bounding box. Although it is not in the scope of this work, it is important also to correctly explain the difference between instance segmentation and semantic segmentation. Both are segmentation methods where each pixel is classified as belonging to different objects, but in instance segmentation, we can separate different objects belonging to the same class, while it is not possible to do in the case of semantic segmentation [158].

In the following sub-sections, some of the algorithms that use CNN for instance segmentation will be presented, where some parts are based in the algorithms for object detection presented previously.

Mask R-CNN

The Mask R-CNN is perhaps the most popular CNN based structure for instance segmentation. This method is based on Faster R-CNN [157], which was briefly explained in the previous sections, but Mask R-CNN has an additional branch to predict segmentation masks on each RoI. In Mask R-CNN an image is run through a CNN to generate the feature maps, which are used by a Region Proposal Network (RPN) to produce multiple RoI using a binary classifier. To achieve high performance, a Feature Pyramid Network (FPN) is used to extract in-network feature hierarchy, where a top-down path with lateral connections is augmented to propagate semantically strong features. The architecture of the Mask R-CNN method is shown in Figure 26.

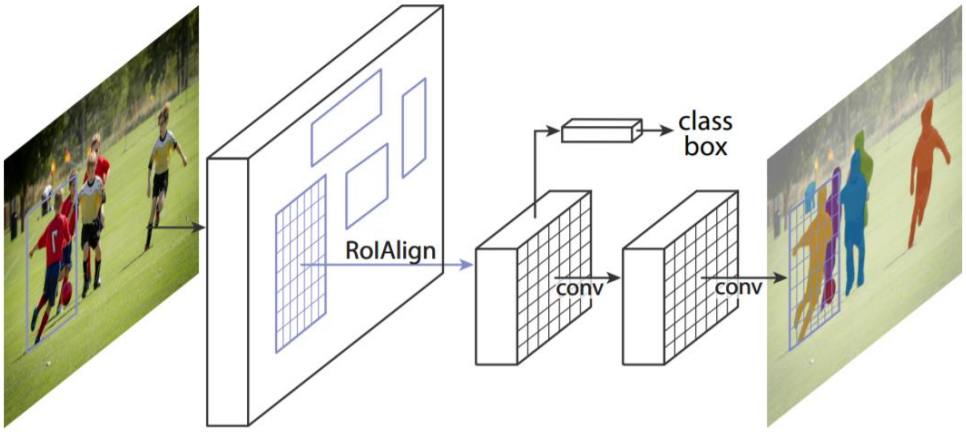


Figure 26. Architecture of the Mask R-CNN method. Adapted from [159]**.**

Practical applications show that this method sometimes can provide high classification scores associated with misaligned masks. This happens since the confidence of instance classification is used as a mask quality score in most instance segmentation frameworks. The mask quality, which is quantified as the Intersection over Union (IoU) between the instance mask and its ground truth, is usually not well correlated with classification score. As masks are predicted by a subnet, specifically conceived for this purpose, in principle mask improvements will be associated with the improvement of this network substructure.

Improvements of Mask R-CNN method

The Mask Scoring R-CNN (MS R-CNN) [160] (Figure 27) is inspired by the idea of directly controlling the mask production inserting a loss for the mask quality (which requires a new network branch named MaskIoU subnet, allowing end-to-end training). The maskIoU subnet uses information from the most likely predicted mask along with information coming from the RoI alignment module, therefore only existing information in the mask subnet is used.

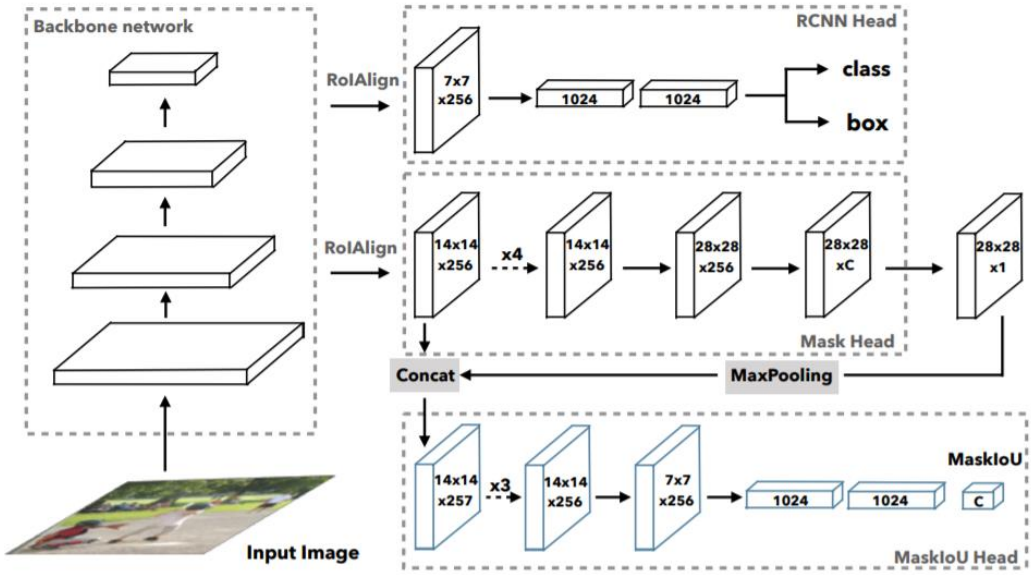


Figure 27. Architecture of the MS R-CNN model. Adapted from [160]**.**

Boundary Mask R-CNN (BMask R-CNN) [161] (Figure 28) is the only that presupposes incomplete information at the mask subnet structure, since information propagation from low levels of the FPN is proposed. This information comes to the mask subnet from an alternate channel and is mixed by the information coming from the RoI alignment module by an intricate scheme along the FCN pipeline.

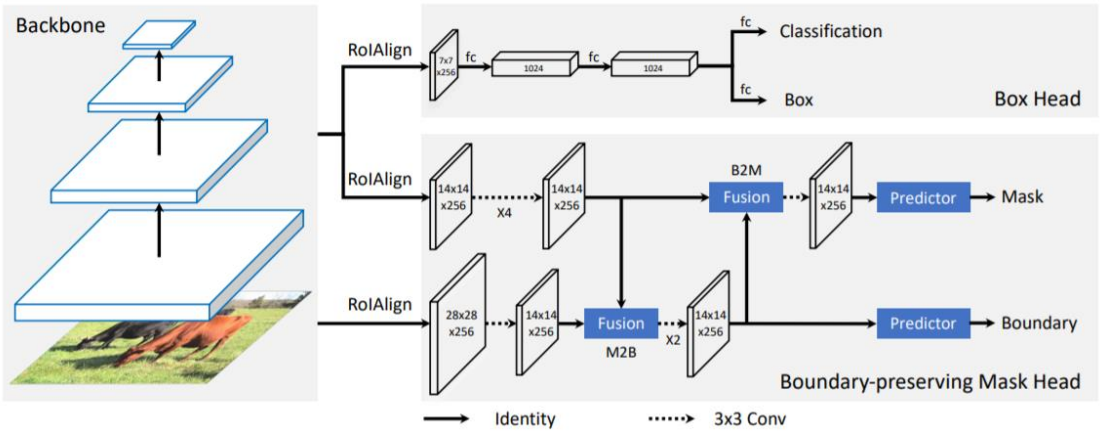


Figure 28. Architecture of the B-Mask R-CNN model. Adapted from [161]**.**

In the Path Aggregation Network (PANet) model [162] (Figure 29), the proposed subnet mask fuses predictions from two views; the conventional FCN (which exists in Mask-RCNN) and small fully-connected layers, which possess complementary properties of FCN. These complementary properties are the core of the method since information diversity is increased, improving the quality of produced masks. Similarly to MS R-CNN, no information from other

modules are propagated to the mask module, while assuming that all the required information for producing better masks exists in the mask subnet. However, this feature can be improved so it can achieve greater performances.

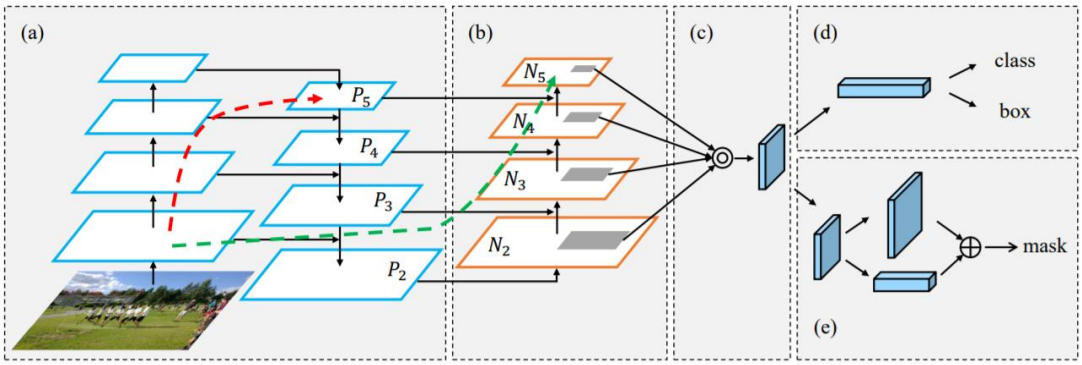


Figure 29. Architecture of the PANet model. Adapted from [162]

4.4.3.4. Mask Improved R-CNN (MI R-CNN)

A novel method for instance segmentation to use in WCE exams is proposed in this section. The Mask Improved R-CNN (MI R-CNN) is an enhanced version of both mask subnets found in Mask R-CNN and PANet models, for multi-pathology detection and lesion localization in WCE videos.

While the classification sub-net of the Mask R-CNN and PANet works well, some misalignments between the predicted and ground truth (GT) masks are found, and therefore, the mask sub-net needs to be improved. As seen previously, MS R-CNN adds a sub-net to the baseline Mask R-CNN to learn the quality of the predicted instance masks. Authors report Average Precision (AP) improvements of more than 1% in the COCO dataset. However, when deepening the results' analysis, it is possible to see that the Mask R-CNN outperforms MS R-CNN for lower APs [160]. This may be related with increased difficulties in the tracking of small objects, since heuristically it seems to be more likely to have higher APs for large objects.

The focus was to improve PANet results by improving mask predictions which is the weak point of the original model. However, regarding to this approach two points must be taken into consideration:

1. Although in [160] a direct comparison between MS R-CNN and PANet using the demanding COCO dataset is not made, when looking at both papers ([162] and [160]), it is possible to conclude that PANet shows improvements in the

performance when applied to COCO dataset. This was the main reason to use the PANet sub-mask module structure in this.

2. It is referred by MS R-CNN authors that there is room for mask prediction improvements in their method, which go from 2.2% to 2.6%, depending on the used backbone [160]. These results were found by changing the predicted mask for the GT in the training process.

The approach of BMask R-CNN [161] is based on boundary information captured from the lower level of the pyramidal structure that enters to the new boundary-preserving mask sub-net, in which object boundary and mask are mutually learned via feature fusion blocks. The method works well, however the data flow in the mask subnet is intricate and complex, and consequently hard to improve. Additionally, authors do not clearly separate and analyze improvements due to information propagation from lower pyramidal levels from improvements due to mask subnet changes. Better improvements are reported in the COCO dataset when compared with MS R-CNN (1.2 and 1.5 % respectively), however both methods present similar behaviors since AP improvements are higher for higher APs. It was also previously proven that for APs lower than 0.5 the improvement of BMask R-CNN over Mask R-CNN becomes negligible, becoming perhaps worsen for smaller APs [161]. These results clearly show poor boundary refinement in small objects.

Instead of trying to add information that comes from different substructures of the Mask-RCNN to the mask sub-net (MS R-CNN) [160], or merging mask sub-net information with information from the RoI alignment sub-net (BMask R-CNN) [161], it was proposed to make a more efficient use of the information that already exists in the mask subnet from the RoI alignment sub-net. This approach is followed in PANet [162], where tiny fully connected layers which hold complementary properties to FCN (used in Mask-RCNN), can capture different views of each proposal. This increases the information diversity, hence producing better quality masks. Several approaches show that propagating low-level information and combining information from different levels makes better use of the entire available information as observed in pyramidal based structures such as the FPN, which is present in both Mask R-CNN and PANet. This approach, that was taken into account at the feature extraction level, was inexplicably ignored at the subnet mask module. The proposed MI R-CNN model improves the quality of predicted masks by propagating low-level information and combining information from different levels in the mask subnet proposed

in PANet [162], which is already an improved version of the mask subnet proposed in Mask R-CNN [159].

In order to preserve most of the information along the CNN pipeline, it was proposed to join in the C4_fc layer information from all the preceding layers of the mask sub-net. The proposed structure is shown in Figure 30 emphasizing the improvement of the Mask Head (mask sub-net) over the approach proposed in PANet [162]. In the first stage (backbone), FPN extracts features to generate RoIs via Region Proposal Network (RPN) for classification and bounding box regression, as it was proposed in Mask R-CNN [159]. The mask sub-net uses each RoI features via RoIAlign, which preserves spatial information, for predicting segmentation.

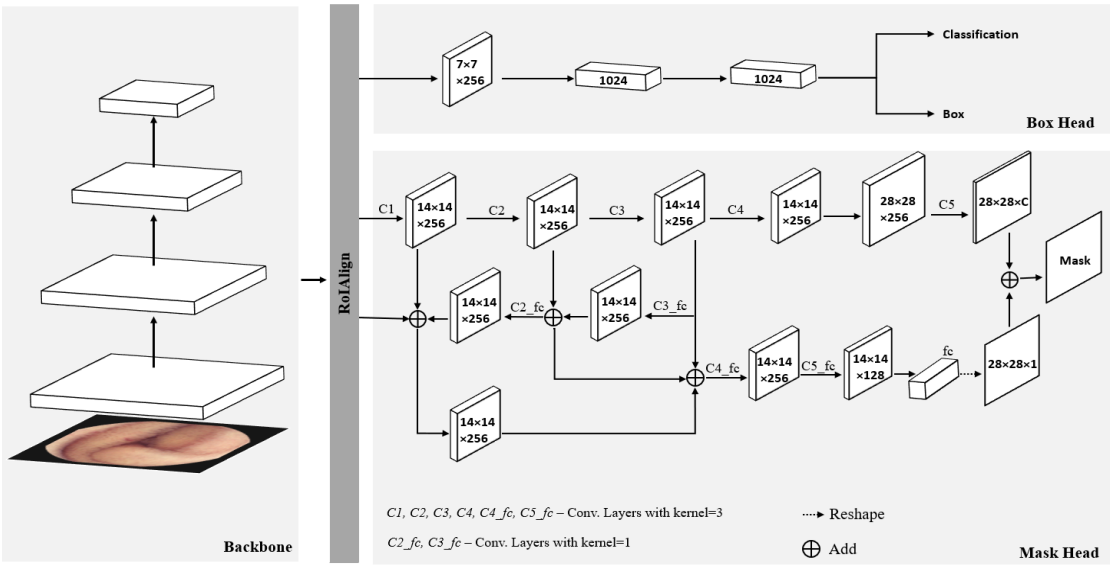


Figure 30. Schematic diagram of the proposed MI R-CNN structure for the mask sub-net proposed in this paper.

4.4.3.5. Model Training

One of the most used methods for Neural Network training is the well-known Stochastic Gradient Descent (SGD) algorithm with momentum. The idea behind the momentum is to take advantage of the convergence dynamics on past iterations to more accurately predict the next one. Momentum codes directly the velocity of the weights' variation along the training process.

The SGD update rule is given by:

$$w^{k+1} = w^k - \eta^k \nabla w^k f(w^k) \tag{33}$$

where f(w) is the function to be minimized, k stands for iteration number and η^k is the learning rate

parameter. The minimization of $f(w)$ can be accelerated by the SGD with momentum method. Momentum is given by:

$$\begin{aligned} z^{k+1} &= \beta^k z^k + \eta^k \nabla w^k f(w^k) \\ w^{k+1} &= w^k - z^{k+1} \end{aligned} \tag{34}$$

where β^k is an iteration dependent parameter. Details on how the η^k and β^k parameters must be updated can be found in [163]. The combination of the pair of equations (34) results in a new update rule given by the momentum:

$$w^{k+1} = w^k - \beta^k z^k - \eta^k \nabla w^k f(w^k) \tag{35}$$

By comparing equations (33) and (35) we can see that the momentum inserted the term $(-\beta^k z^k)$ in the update rule. By using the last equation of the pair of equations (34), equation (35) can be rewritten as:

$$w^{k+1} = w^k + \beta^k (w^k - w^{k-1}) - \eta^k \nabla w^k f(w^k) \tag{36}$$

Therefore, the term inserted by the momentum is the first difference of the network weights weighted by the β^k parameter that must be adjusted.

Nesterov Accelerated Gradient (NAG) [164] computes equation (36) on the basis of the estimate of the next position of the parameters instead of on the current position. In addition to the momentum, NAG also significantly accelerates the algorithm convergence and the updating rule becomes:

$$w^{k+1} = w^k + \beta^k (w^k - w^{k-1}) - \eta^k \nabla w^k f(w^k - \beta^k z^k) \tag{37}$$

The acceleration of convergence through the momentum was based on the hypothesis that the successive aggregation of past gradient information is more effective than the latest negative gradient alone. In fact, the step taken at the previous iterate w^{k-1} was based on negative gradient information at that iteration, along with the search direction from the iteration prior to that one, w^{k-2}. By following this line of reasoning, we see that the previous step is a linear combination of all the gradient information found at all iterates so far, going back to the initial iterate w^0.

Higher-order momentum

Gradient information of past iterates is given by the derivative of the weight parameters, which presupposes that at each iteration acceleration coefficients (second derivative of the weight

parameters) can also encode convergence dynamics more extent in time and can help to improve convergence. A similar reasoning can be made regarding higher than second order momentum. As the n^{th} order derivative is just the derivative of the $(n\text{-}1)^{th}$ derivative then the nth momentum is the momentum of the $(n\text{-}1)^{th}$ momentum. Therefore the second momentum, which is the acceleration of the weight coefficients can be obtained by the momentum of the momentum given in equation (34) and is given in equation (38) already with the inclusion of the NAG which was named as NAG2:

$$\begin{aligned} \mathrm{s}^{\mathrm{k}+1} &= \gamma^k s^k + \eta^k \nabla w^k f(w^k - \beta^k z^k - \gamma^k s^k) \\ z^{k+1} &= \beta^k z^k + \mathrm{s}^{\mathrm{k}+1} \\ w^{k+1} &= w^k - \mathrm{z}^{\mathrm{k}+1} \end{aligned} \tag{38}$$

The updating rule becomes:

$$\begin{aligned} w^{k+1} = w^k + \beta^k\left(\mathrm{w}^{\mathrm{k}} - \mathrm{w}^{\mathrm{k}-1}\right) + \gamma^k \mathrm{s}^{\mathrm{k}} \\ - \eta^k \nabla w^k f(w^k - \beta^k z^k - \gamma^k s^k) \end{aligned} \tag{39}$$

From the last two equations of the set of equations (6) we obtain:

$$\gamma^k s^k = \gamma^k [w^k - (1 + \beta^k) w^{k-1} + \beta^k w^{k-2}] \tag{40}$$

Equation (8) shows that the second momentum also reinforce indirectly the first momentum which can be seen rewriting equation (39) by inserting equation (40). Equation (39) becomes:

$$\begin{aligned} w^{k+1} = w^k + (\beta^k + \gamma^k)\left(\mathrm{w}^{\mathrm{k}} - \mathrm{w}^{\mathrm{k}-1}\right) - \beta^k \gamma^k \left(\mathrm{w}^{\mathrm{k}-1} - \mathrm{w}^{\mathrm{k}-2}\right) \\ - \eta^k \nabla w^k f(w^k - \beta^k z^k - \gamma^k s^k) \end{aligned} \tag{41}$$

For a more detailed analysis of the convergence of the second momentum and some results around it, it is suggested to consult the reference [163].

4.5. Conclusion

This chapter was dedicated to the presentation of different methods used along the development of this thesis. First, a brief overview about color spaces and how the different channels have different representation of the lesions found in the GI in WCE exams, along with the presentation of some different features that can represent the different abnormalities.

Next, the implemented segmentation procedure was presented, with a special focus on acceleration methods and MRF for neighboring improvement.

Finally, the different classification models and schemes used in the ambit of this work were presented. First, single learning models were addressed, specifically MLPs and SVMs. These are considered one of the most basic classification models in Machine Learning theory, and are usually used for simple tasks when a small number of features are present. Afterwards, ensemble learning models are described, more specifically the proposed scheme used in this work. In the end, Deep Learning models are also included in the ambit of analysis; with a focus on the conventional CNN structure, the networks used for object detection and instance segmentation, their drawbacks, and the proposed network for this topic (MI R-CNN).

.

Chapter 5. Angioectasia Detection

This chapter describes the work done regarding angioectasia detection in WCE videos. Since this work was published in [6] and [5], this chapter is heavily based on those two papers. While in the first paper only angioectasia segmentation with EM algorithm with MRF was addressed, in the second paper a weighted-boundary function was included in the MRF and also classification methods to detect these lesions.

5.1. Overall Methodology

Figure 31 shows the flowchart of the proposed system, which consists in two major blocks: segmentation and classification modules. The purpose of the segmentation module is to break the image into two logical regions, where each logical region may have several space regions since lesions may be scattered across the frame. The purpose of the classification module is to characterize the lesion nature by extracting adequate features in both logical regions. Lesion nature knowledge was previously acquired in the training phase. Angioectasias seem to have a better characterization in the CIELab than in the RGB color spaces, as discussed in Section 4.1.2. Therefore, CIELab will be the color space chosen in the current approach. A new acceleration

procedure for the EM algorithm is used (Section 4.3.2.) in order to avoid slow convergence especially in normal frames where the classes may be poorly separated. An MRF approach allows to compensate the assumption of independence of the GMM model among different components, by adding spatial information to the segmentation procedure. A relaxation coefficient computed based on pixel intensities is proposed in order to cope with boundaries' imperfections (Section 4.3.3.). A sigmoid-based function is proposed to compute this coefficient, which was obtained by experimentation. Regarding feature extraction for classification purposes several statistical measures are computed in both regions where one is considered background (normal tissue). This procedure leads the system to focus more on the differences between regions than in absolute values in each region, improving perhaps robustness against patient and device variability being a significant novelty of the proposed approach.

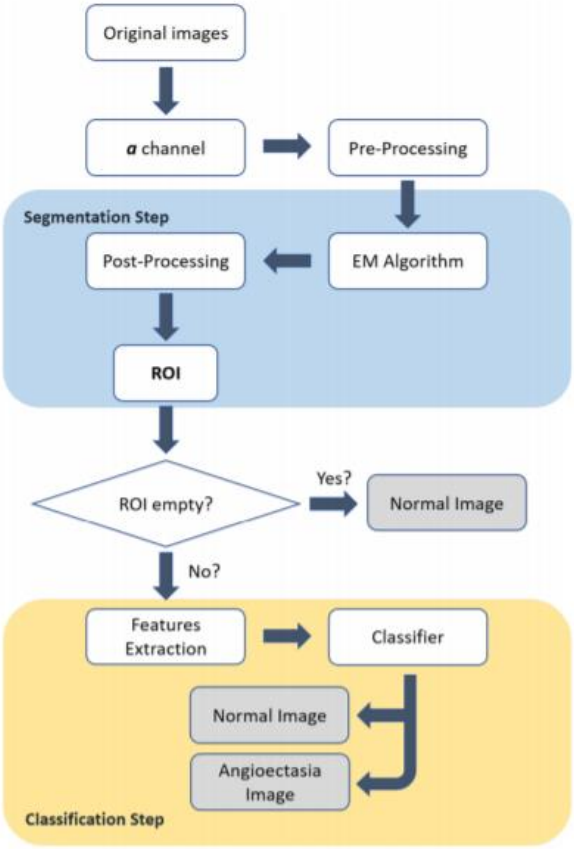


Figure 31. Pipeline of the implementation of the angioectasia detection.

5.2. Results

This section presents the used datasets and some implementation details, followed by the results focused on the segmentation module and ending with global classification results.

5.2.1. Dataset and Implementation

In this work, two different datasets were used:

1. For the evaluation of the segmentation module, the public database KID was used [165]–[168]. This database consists of 27 images, divided into 3 groups of different bleeding probabilities (P0, P1 and P2; from the lowest probability to the highest). All the images were manually segmented by experienced physicians and were all acquired with MiroCam®.
2. For detection purposes, a bigger database was used, with 798 images (248 images with angioectasias and 550 images labeled as normal). All the images with lesions and 300 normal frames were taken from 20 exams from PillCam™ SB2, where the rest of normal images (250) were taken from 5 normal exams from MiroCam® in order to obtain a higher degree of generalization. All the exams were performed in Hospital of Braga (Portugal) and were examined by two expert physicians in the diagnosis of WCE exams. The images were included in the database only when both agreed with the diagnosis.

Experimental results of classification were obtained by using WEKA – an open source machine learning package. A stratified 10-fold cross-validation was used, taking into consideration subject variability along folds, with a MLP neural network and a SVM. For evaluation purposes, several metrics were computed for each test: sensitivity, specificity and accuracy, computed as follows, as well as area under the ROC curve (AUC):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{42}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{43}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{44}$$

A baseline system was implemented, with the purpose of comparing the proposed method with a reference result. This implementation was based on the algorithm described in [82]. This method applies a histogram equalization step to increase contrast of the images and a decorrelation between RGB channels to enhance color differences. After, a threshold is applied to the green channel, which will work as a seed to a region growing algorithm. Regions with specific values of area, perimeter and extent are then removed; and the rest of regions are splitted when their variance were higher than a specific value. Then, 24 statistical, textural and geometrical features are extracted for several color spaces (RGB, HSV, CIELab and YCbCr). A decision tree (RUSBoosted) is used to classify each region as normal or abnormal.

5.2.2. Color Space and Pre-processing

The purpose of this sub-section is to show that the CIELab Color Space has some advantages over RGB even if the relative Red value is used, as defined in Eq. (45). For this purpose Figure 32 presents a frame containing an angioectasia shown in different color channels of different color spaces. It is clear that the *a* channel (Figure 32d) clearly shows the lesion, while some potential false positives are avoided. R_* channel (Figure 32c) also avoids potential false positives however the lesion appears much more subtle. Since RGB shows a high correlation among the three channels [64], the relationship between color red and channel R is not direct. In other words, information carried by the red component is also carried by the other two components, hence a large range of information is simultaneously carried by the different color components. Therefore, even using the R_* component, the discrimination remains difficult, however slightly better when compared to the red component (R).

$$\mathrm{R}_* = \frac{\mathrm{R}}{\mathrm{R} + \mathrm{G} + \mathrm{B}} \tag{45}$$

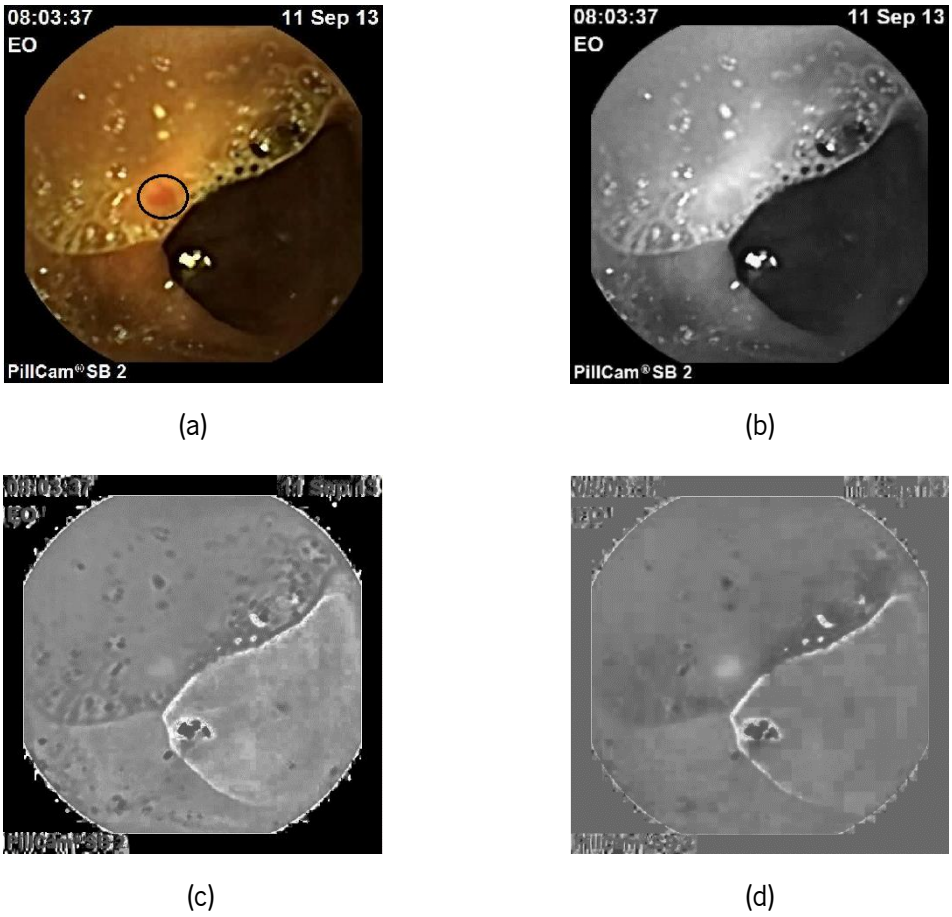
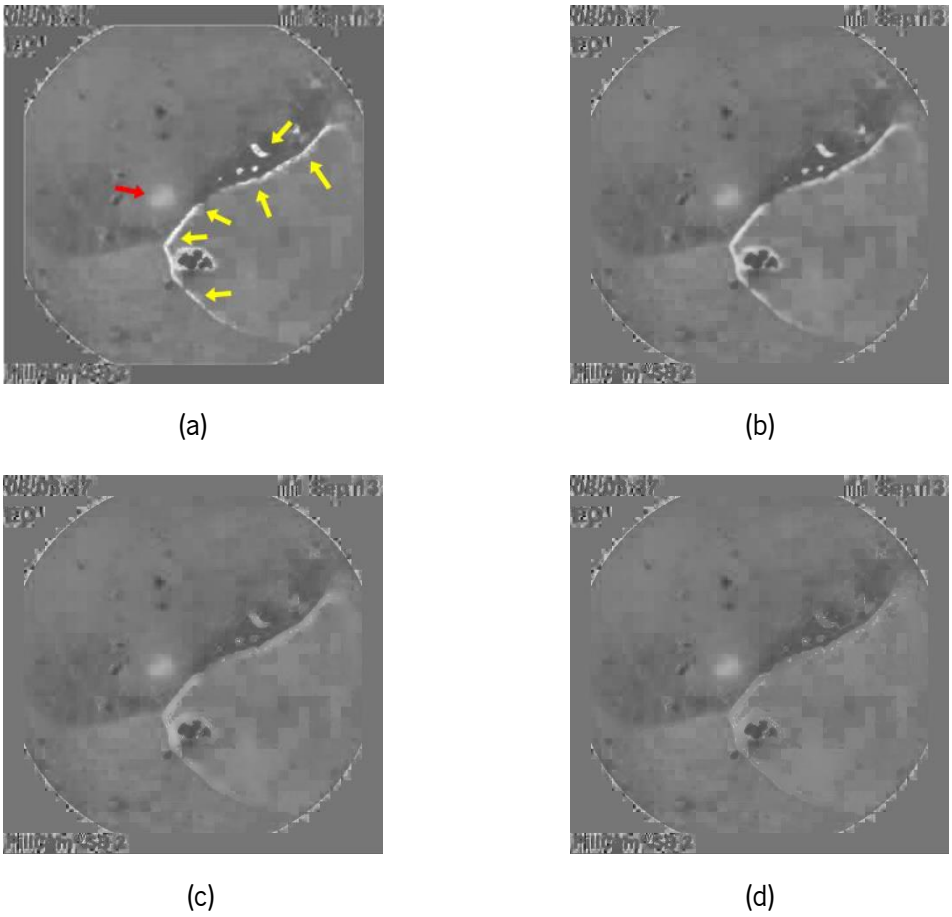


Figure 32. Example of an angioectasia in the SB taken from a WCE exam (a), red component of image in RGB (b), relative red component in RGB (c) and *a* channel from CIELab (d).

Figure 33 shows some results after using the preprocessing step, previously described in Section 4.1.3. where an angioectasia frame is used as an input with different neighborhood sizes (7, 21 and 51). Results show that with an increasing size of $\mathcal{N}\{D^l(i,j)\}$, the algorithm shows a better performance. Although the lesion also becomes less intense when increasing the neighborhood size, this pre-processing step improves the overall results of segmentation (as will be discussed in the next subsection). The chosen δ was 5, because presented good results for removing these highlighted regions, not affecting the lesion area.



(a) (b)

(c) (d)

Figure 33. Original a component of CIELab space color (a) and the same component with the pre-processing step with a neighborood of 7 pixels (b), 21 pixels (c) and 51 pixels (d).

5.2.3. Segmentation

The segmentation of angioectasias is an important module of the whole system, because it strongly influences the next step (Classification). The segmentation algorithm proposed in this paper is an improvement of the one presented in Reference 30, which was also tested in the KID dataset. Major improvements are the acceleration of convergence of the EM algorithm and a new parameter b_k which incorporates pixel intensity in the computation of *a priori* probabilities provided by the MRF.

Figure 34 shows segmentation results of an image with an angioectasia lesion. Figure 34b shows the performance of the Otsu algorithm applied to the *a* channel of the frame. It is clear that with this basic segmentation method it is impossible to correctly differentiate normal from angioectasia tissue. Figure 34c shows that the use of MAP algorithm leads to a major improvement when compared to the Otsu method. The same argument can be used when looking at Figure 34d

and Figure 34e; where both pre- and post-processing methods lead to an improvement in the segmentation of this image. In the case of the pre-processing step, almost all non-lesion zones highlighted with *a* channel were removed. With the use of post-processing, small regions were also removed, reaching a result where the only selected pixels are the ones belonging to the angioectasia.

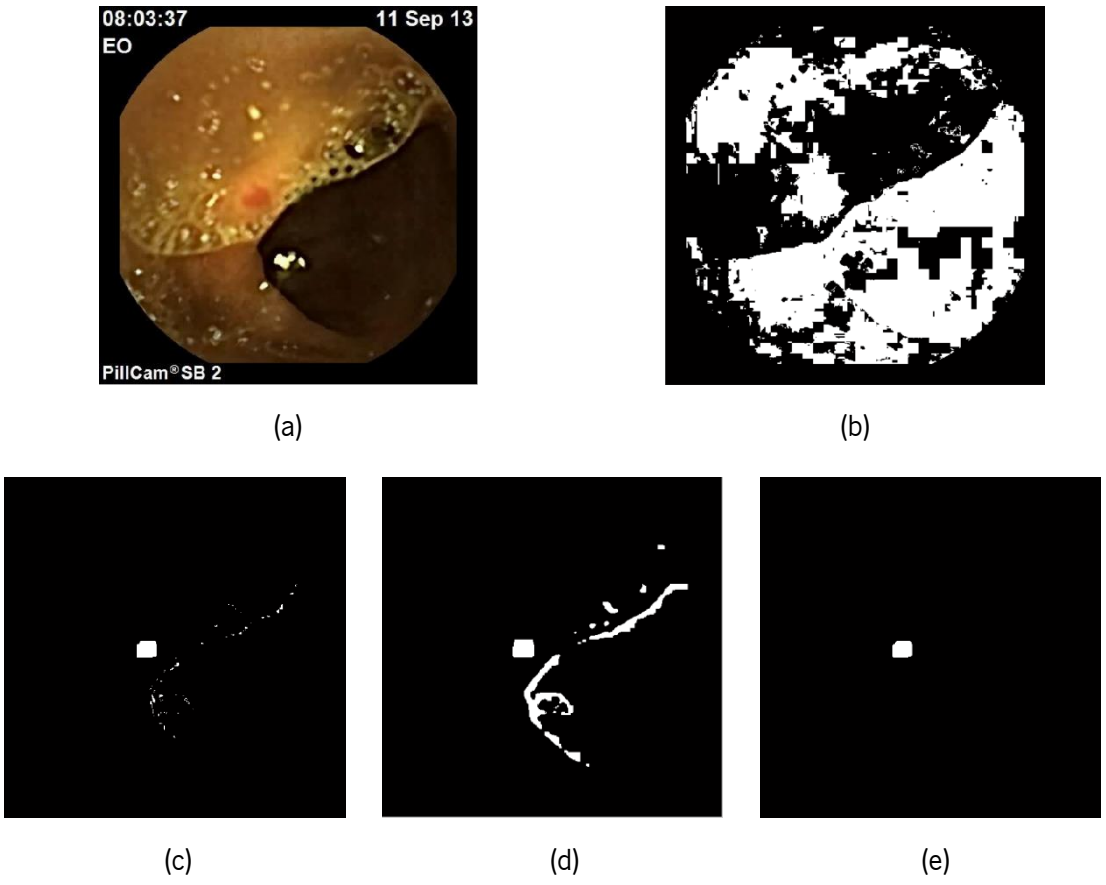


Figure 34. Image with an angioectasia (a), segmentation results with Otsu thresholding of component a (b), MAP without preprocessing (c), MAP with pre-processing but no post-processing (d) and MAP with pre- and post-processing (e).

Figure 35 shows another example of a frame with an angioectasia, but in this case there is a higher incidence of bubbles in the image. This fact increases the number of pixels corresponding to reflections in these bubbles, that contain a high value of the *a* component (which can be verified with the analysis of Figure 35b). When no preprocessing is applied, the region of angioectasia is not selected by the segmentation module (Figure 35c). The result is improved with the inclusion of the pre-processing step, in which the angioectasia lesion is also selected (Figure 35d). This result is also improved with the inclusion of the post-processing step (Figure 35e), exactly as in the previous example.

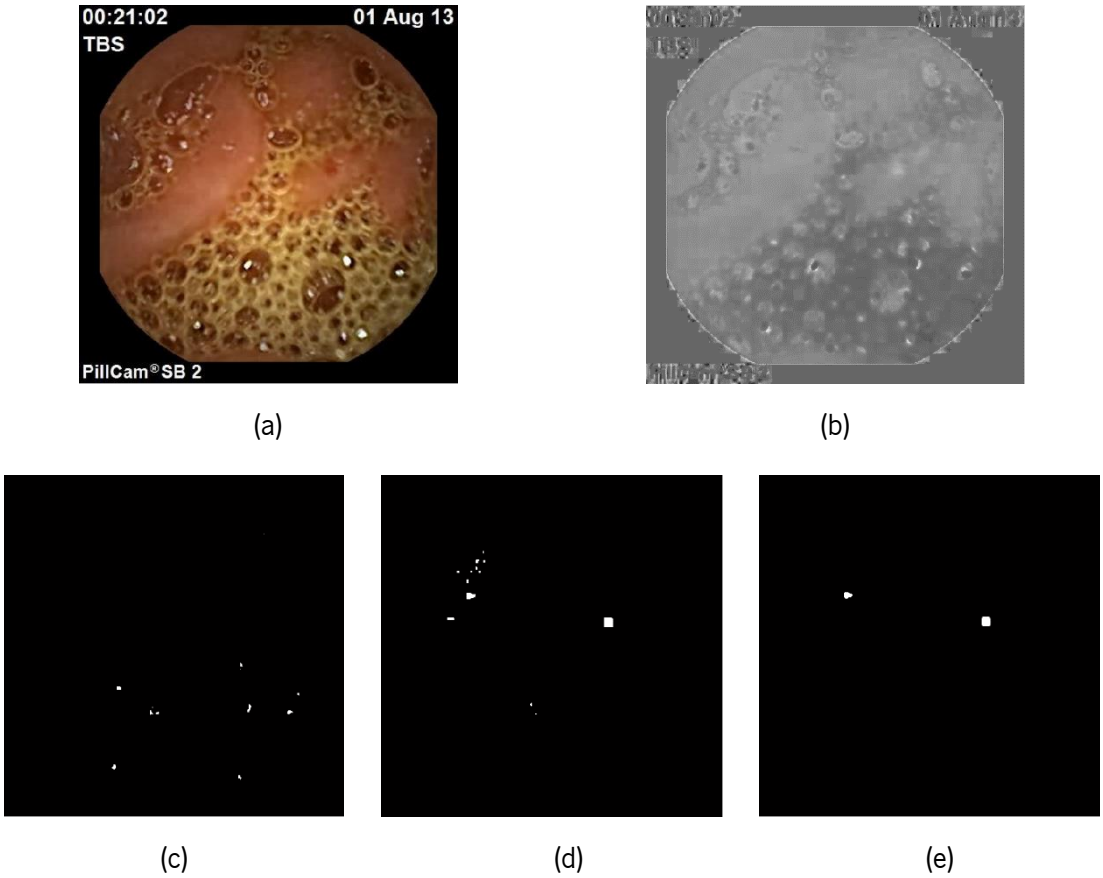


Figure 35. Image with bubbles and an angioectasia (a), channel a of the same frame (b), MAP without pre-processing (c), MAP with pre-processing but no post-processing (d) and MAP with pre- and post-processing (e).

KID Database was used to validate the segmentation algorithm since manual segmentation of all the images are available. To compare the different methods the Dice metric (Eq. (46)) was computed over all the 27 images of the database. In this equation, A is the set of pixels segmented by the algorithm and B is the set of pixels in the manual segmentation. The higher the Dice metric, the better is the performance of the segmentation algorithm.

$$\mathrm{D(A,B)} = \frac{2|\mathrm{A\backslash capB}|}{|\mathrm{A}| + |\mathrm{B}|} \tag{46}$$

To see how every step would influence the segmentation of the images 4 different experiments were carried out:

1. Exp A Results without pre and post-processing, using MRF with constant β values.
2. Exp B Results with pre-processing but without post-processing, using MRF with constant β values.
3. Exp C Results with pre and post-processing, using MRF with constant β values.

4. Exp D Results with pre and post-processing, using MRF with varying β values.

Figure 36 shows Dice values for both the whole KID dataset (a) and only P2 lesions from the same dataset (b). These lesions are the ones with the biggest probability to bleed, therefore they are most dangerous for the patients. Also, it is common for these lesions to appear bigger and more reddish.

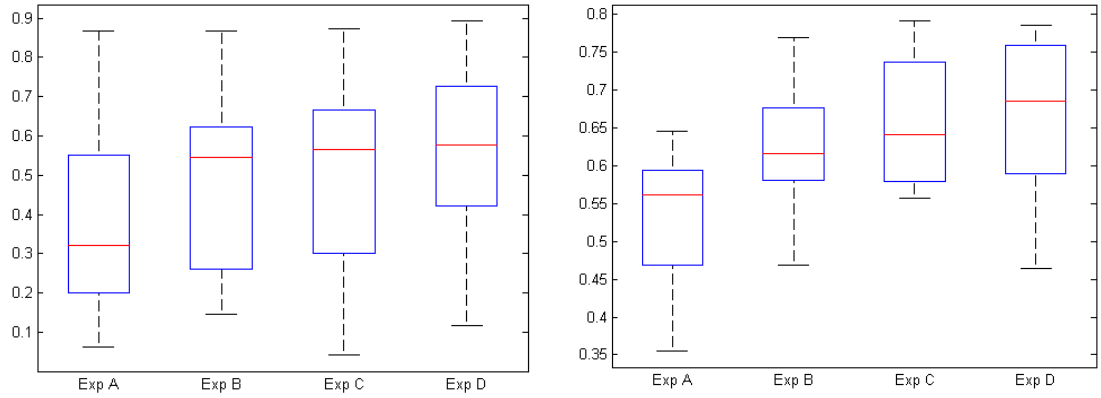


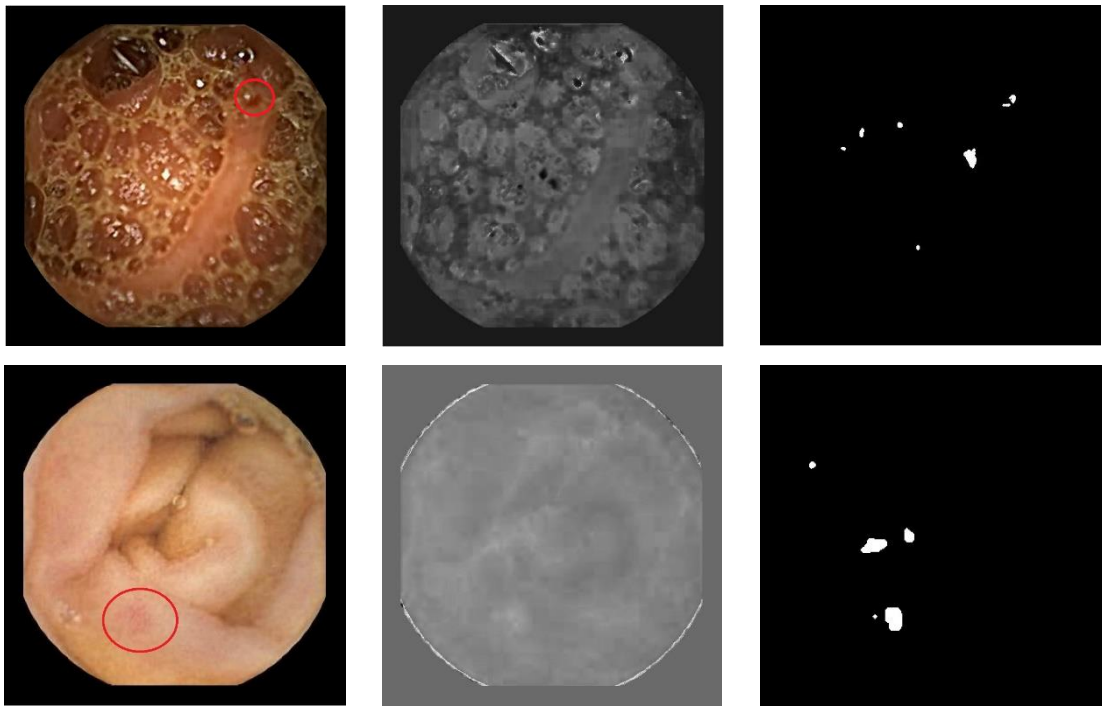
Figure 36. Box plots of Dice values after 4 different experiments with the whole KID Database (a) and only with P2 lesions (b). In each plot, from left to right, without pre and post processing, only with pre-processing, MRF with constant β values and MRF with varying β values.

Results allow to infer the efficiency of each block of the segmentation algorithm; pre-processing, post-processing and intensity based anisotropic MRF modeling. The improvement is more prominent from Exp. A to Exp. B, showing that indeed the pre-processing step has a relevant role in the segmentation method as a whole. The inclusion of post-processing also leads to an improvement in the segmentation performance (Exp. B to Exp. C). Lastly, an improvement is also notorious when comparing standard MRF with the new approach proposed in this paper (Exp. C to Exp. D).

When comparing the results from Figure 36a and Figure 36b, the more immediate conclusion is that Dice values are bigger when only using P2 lesions, which was concluded in a previous paper [6]. It is also important to notice that P2 lesions show a bigger improvement when the proposed equation for MRF is included. This fact shows that this new parameter is important for segmentation of this type of lesions that have sharper edges.

It can be also interesting to analyze some images that show some flaws in the Segmentation step, which will directly affect their classification afterwards. In Figure 37, both images led to bad segmentation results, as can be seen in the images on the right. The first

example (top), the lesion is behind the bubbles, which becomes difficult for the system to localize it (channel a is highlighted in lesion tissue and in presence of bubbles). Although the system could find some of the angioectasia tissue, most part of the result is not correct. In the bottom example, the angioectasia is at an early stage, appearing with a very soft red appearance in the middle of the tissue. As it can be seen in Figure 37b, abnormal tissue is not that highlighted, which leads to the result on the right where several false positives appear.



(a) (b) (c)

Figure 37. Examples of misclassified frames (angioectasias with red circles) (a), with the respective images representing the channel *a* (b) and the segmentation result after applying the proposed method (c).

5.2.4. Classification

One of the advantages of the approach presented in this paper is that not all images need to be passed in the classifier to be classified. Some of the normal ones are classified by the segmentation module because they do not have RoI. In fact, normal frames present smooth amplitude variations between neighboring pixels, which causes the most intense class to be spread in numerous groups of few pixels that are easily eliminated in the post-processing step. Table 2 shows that more than 33% of normal frames were correctly classified as normal and none of the

pathological ones was classified as normal. This is a major advantage for the classifier enabling a more effective discrimination by reducing the sample space.

Table 2. Results after the first step of the algorithm.

Type of images	Normal	Angioectasia
# of images in the Database	550	248
# of images classified as normal after segmentation step	182	0
# of images used in the classification step	368	248

Given the findings and conclusions of Section 4.2 ''Extraction of Features'', second and higher order statistics were used in the ambit of this paper. A feature analysis algorithm was used to rank the different features according to its discriminative and predictive power. A correlation-based feature selection, which computes the Pearson's correlation value between the value and its class, was used. These results can be analyzed in Table 3, where all the features were ranked according to the correlation values obtained using WEKA.

Table 3. Correlation based feature selection.

Feature	Correlation value
μ_a	0.7845
σ_a^2	0.6701
H_b	0.6590
H_a	0.6587
μ_b	0.3647
k_a	0.1640
σ_b^2	0.1109
k_b	0.0045

It is clear that *a* channel presents the most representative features, where the mean reaches the highest value. It was decided to group features into different sets, which were built according to the results in the previous table and according to the statistical measures that were expected to represent in a better way the data. The chosen sets can be seen in Table 4.

Table 4. Sets of features tested in angioectasia detection.

Set	Features
A	μ, σ^2
B	μ, H
C	μ, k
D	μ, σ^2, H
E	μ, σ^2, k
F	μ, σ^2, H, k
G	4 most representative features according to Table 3

In Table 5 and Table 6 were included the results with an MLP classifier and a SVM classifier, respectively. Not surprisingly, given the non-Gaussian nature of the intensity distribution, the best results were achieved by including higher order statistics. It was also observed that entropy seems to have more discriminative power than kurtosis. Results present higher sensitivity than specificity, which is usually the purpose with medical applications, because it is more acceptable to have some false positives than false negatives.

Table 5. Results using a MLP classifier and different sets of color features from *a* and *b* components of Lab color space.

Set	Sensitivity	Specificity	Accuracy	AUC
A	93.52	89.34	91.86	90.21
B	91.24	86.91	89.53	91.39
C	92.35	91.07	91.86	88.98
D	94.33	89.30	92.33	93.46
E	94.69	92.94	93.95	94.51
F	96.60	94.08	95.58	94.98
G	92.41	91.72	92.19	93.42

Table 6. Results using a SVM classifier and different sets of color features from *a* and *b* components of Lab color space.

Set	Sensitivity	Specificity	Accuracy	AUC
A	96.57	89.93	93.95	88.23
B	93.53	76.73	86.98	89.34
C	93.92	88.75	91.86	92.12
D	96.20	88.71	93.26	93.56
E	95.04	91.69	93.72	93.89
F	96.58	92.24	94.88	95.62
G	94.36	86.05	91.65	90.21

Another interesting finding is that, when using the features that were chosen to be more representative according to Table 4, the results did not show an improvement. For example, the set C (which is a set with also four features), presents similar results both when using MLP or SVM. When using SVM, set A is the set with only four features with the best performance. Both tables also show that MLP and SVM perform similarly (mainly in accuracy), where the sensitivity values are higher in SVM and specificity lower. The best result was achieved using an MLP classifier (96.60% of sensitivity and 94.08% of specificity, leading to an accuracy value of 95.58%).

When comparing with the reference algorithm implemented by the authors (Table 7), the proposed approach led to higher results both in sensitivity and specificity. The selected method [82] is based on several parameters chosen by its authors, when using a specific database, which is why the results shown in here are not equivalent to those shown in the paper. We can conclude that the performance of the proposed overall system is better than the one considered as a reference.

Table 7. Results of the reference algorithm based on the algorithm explained in Reference [82] and the best results of the current approach (considering the classification of both steps).

Set	Sensitivity	Specificity	Accuracy
Results of reference [82]	87.50	86.40	86.79
Current approach	96.60	96.55	96.58

5.3. Conclusion

Lesion detection in WCE exams is of a tremendous interest for the future of medicine, and the development of an automatic diagnosis system that deals with this problem will allow for the physicians to reduce the reviewing time of these exams. This work presents an effective approach for segmentation and detection of angioectasias in SB tissue in WCE exams.

With this work, it was possible to conclude that the use of CIELab color space indeed improves the highlight of angioectasia lesions in WCE frames, specifically when the proposed pre-processing algorithm is applied. The use of MRFs to model the pixels' neighborhood also shows improvements in lesion segmentation, more specifically with the inclusion of the proposed weighted-boundary function. Both of these improvements were shown visually (Figure 34 and Figure 35) and graphically (Figure 36). Another important novelty presented in this paper is that the segmentation module has the ability of correctly classify a substantial percentage of normal images of the dataset. This is important to reduce the time needed to classify an entire WCE video. The use of a supervised classifier showed that these lesions can be detected with the use of a small set of features. As expected, higher order statistics improve the system performance, given the non-Gaussianity observed in intensity distributions.

When compared to the methods already published to detect angioectasias in WCE images, the proposed system shows a better performance. This method does not use algorithms with a high computational complexity (like deep learning), which can be an advantage to be used in a clinical practice, where machines with a high computational power are not usually available. Also, the work described in this paper does not just has the ability of detect angioectasia lesions, but also can localize them in the image; which can be an extra help for the physician.

Nevertheless, there are still some problems to address. As was shown in Figure 37, some images were misclassified with the proposed method. The pre-processing step should be improved so images with different backgrounds would have the lesions better highlighted. Also, in future works, maybe other color channels could be used simultaneously to segment angioectasias. Also, the segmentation has room for further improvement (specially when smaller lesions are present), which consequently will improve the classification. More promising features to encode color information should be tested, like histogram of oriented gradients (HOG). And improved classification methods should be used (like deep learning or ensemble learning), that usually need larger databases in order to work properly. Looking at the performance values that were reached, we can conclude that the use of this system in the clinical practice can be started, as well as tests with entire videos of WCE and the test of this system in clinical practice.

Chapter 6. Tumor Detection

This chapter describes the work that led to the development of a system for automatic tumor detection. This work was published in [7] and this chapter is extensively based on this paper. This publication was preceded by [169], where some tests for both segmentation and classification were described.

6.1. Overall Methodology

The proposed approach tries to code texture in the image domain in order to save computational resources. The main idea is to characterize tumors with an efficient color code that does not lose performance. To mimic the changes that physicians search for, a perceptually based color space will be used; and to deal with the light variations among images, a color space with the luminance separated from the color components will be convenient. In fact, having direct access to the luminance, can help to cope with light saturated regions, since light intensity variations change color perception. Therefore, the CIELab color space seems to be appropriate for this application since lightness appears separated from both color-opponent dimensions and apart from that, this color space is perceptually motivated. Regarding features extraction, the best

performance was obtained by using histogram-based measures, for both color-opponent dimensions. It is well accepted by the scientific community that using an ensemble of classifiers (each one specialized in a different subset of the training data), leads to a better performance than when using a single classifier [144].This work proposes the use of the EL concept based on SVM by using a modified version of the on-line Bagging strategy, where the training data is partitioned by a multivariate GMM instead of using bootstrap. The usual method of the simple average for combining models' outputs is changed for a nonlinear combination, accomplished by a conventional MLP. This scheme is usually known as a stack of classifiers. While usually all the classifiers in the stack are trained simultaneously this paper proposes a new training strategy, in which the goal is to preserve diversity in incremental (online) adaptation. The pipeline of the proposed method is shown in Figure 38.

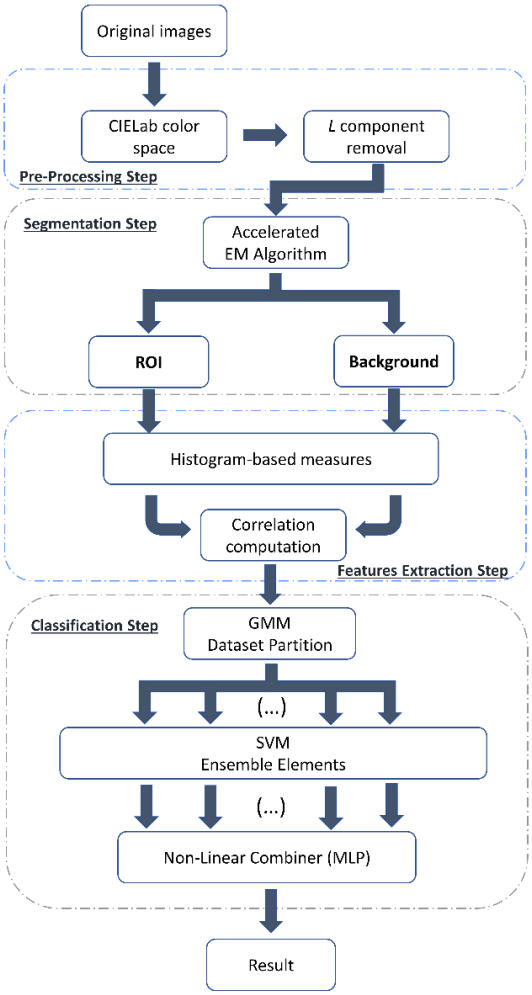


Figure 38. Pipeline of the proposed system for tumor detection.

6.2. Results

The main purpose of this section is to show the different results that have been computed so that the effectiveness of the proposed method can be evaluated. First, some considerations are made about the implementation and description of the used dataset. Afterwards, a joint analysis of segmentation, features and classification methods is presented.

6.2.1. Dataset and Implementation

A 2.3 GHz Intel i7 quadcore processor, with 6 GB of RAM, was used with MATLAB to run the proposed algorithm. The average processing time is 0.06 seconds per frame, which leads to a saving in time processing of at least 25 times regarding color correlation texture analysis that requires the computation of co-occurrence matrices, and wavelet or curvelet computations in the three color channels.

For the ensemble classifier, the training dataset is partitioned into four groups in a bi-dimensional space (pathological versus normal) by using a multivariate GMM and the EM algorithm. Each one of the 16 (4X4) ensemble elements (SVMs) is then trained on each partition of the training dataset by using WEKA – an open source machine learning package. A stratified 10-fold cross-validation was used for this purpose. The entire training set is then used for training the combiner element that is a three-layered MLP with three neurons in the middle layer. The WEKA package is also used for this purpose by using again a stratified 10-fold cross-validation.

The limited incidence of tumor in the small intestine and the lack of a standard database for this pathology limit the amount of available data for experimental purposes. Our current dataset contains 936 frames from 29 patients labeled as tumoral frames. These frames were selected by a team of expert gastroenterologists from Capucho's Hospital in Lisbon under the criteria of medical unambiguity, which in some cases required invasive complementary examination. Concerning the type of tumors, from this group of patients, 10 were diagnosed with an adenocarcinoma (359 images), 8 with an intestinal lymphoma (262 images), 7 with a carcinoid tumor (206 images) and 4 with sarcoma (109 images). Regarding normal frames, only 3000 were used in order not to unbalance too much the amount of data belonging to each class. Roughly twenty percent of these normal frames belong to the 29 patients (20 frames from each one) while roughly 80% of this data belong to 10 normal subjects. Given the relatively low number of patients,

the cross-validation strategy was patient-based, instead of the most common approach of individual frames based; in order to obtain a better generalization.

6.2.2. Considerations of the Analysis

The analysis of the results includes not only the characterization of tumoral tissue based on *a* and *b* color-opponent dimensions of the CIELab color space, but also the RoI selection based on EM segmentation and the classification scheme based on an ensemble of SVMs.

In this work, three different aspects were considered:

1. The effectiveness of the proposed features: Color features taken from CIELab color space and state-of-the-art DWT and Discrete Curvelet Transform (DCT) will be extracted. This will help understanding how the new features could encode the tumoral tissue information when compared to the state-of-the-art results.
2. The effectiveness of the RoI selection: In this regard we need to compare the performance of the system using both the full image or the separation of each image into two regions.
3. The assessment of the classifier performance: The ensemble classifier performance will be approached by testing the different sets of features with both situations. First, a state-of-the-art mixture of experts where the elements of the ensemble (SVM's) and the gating network are trained simultaneously will be used and then the proposed ensemble of SVMs.

In this experiment the dataset is imbalanced (the number of tumor frames is smaller than the number of normal frames). This type of data usually induces high accuracy immediately, because the chosen model cleverly decides that the best thing to do is to always predict the class that allows a high accuracy. Consequently, more reliable measures can be used such as sensitivity (true positive rate), specificity (true negative rate) and accuracy [170]. The classification performance is high when both sensitivity and specificity are high.

The proposed features were grouped into five different sets as shown in Table 8. Set X represents all the features used in this work, which results are only shown in the last subsection because in the previous ones no significant improvements were found.

Table 8. Organization of the subsets of features.

Set	Features
A	μ, σ^2
B	μ, H
C	μ, k
D	μ, L_{95}
X	All Features

6.2.3. Analysis of the proposed features

Regarding the analysis of the statistical features computed from the histogram, no segmentation is performed in such a way that the only changes come from the different types of features. The classification is performed by using an SVM trained using the WEKA package with a stratified 10-fold cross-validation.

Results with the proposed features can be seen in Table 9. State-of-the-art texture features taken from both DWT and DCT were also computed for comparison (Table 10). These algorithms have shown very satisfactory performances, however demanding high computational resources. Conceptually these methods and the method proposed in this work differ in one fundamental aspect; while the first relies on the processing of the whole frame the current method rely on an automatic segmentation which purpose is to select tumoral tissue. We argue that if the algorithm can successfully separate (detect) tumor tissue from normal tissue (including non-tissue regions) for segmentation purposes then tumoral tissue can perhaps also be well characterized for classification purposes by using the same reasoning.

The statistical measures chosen in this case were only the mean and variance of the different directions, which proved to represent well the texture of a tissue [171]. In the DWT case only the first scale was considered since the most part of texture information is contained in the high frequency components of the image. For DCT, 16 directions were used also in the first scale.

Table 9. Proposed features taken from the entire image using SVM.

Set	Sensitivity	Specificity	Accuracy	AUC
A	91.3	95.1	94.2	92.1
B	88.5	91.0	90.4	91.3
C	86.0	87.9	87.5	82.4
D	90.5	94.2	93.3	90.3

Table 10. Textural features taken from the entire image using SVM.

Set	Sensitivity	Specificity	Accuracy	AUC
DWT	81.4	91.7	89.3	85.2
DCT	75.4	95.0	90.3	86.7

When looking at Table 9 and Table 10 a clear superiority of the *ab* color subspace when compared with the textural based features is shown, especially for set A that outperforms the other 3 cases. DCT performs better than DWT which is not surprising since DCT considers more directions, perhaps gathers more information of texture in its coefficients. Regarding accuracy DCT outperforms set C and has similar results to set B.

Table 9 and Table 10 also show that although the proposed features present high values for both sensitivity and specificity, textural features have a clear decrease in sensitivity values. A curious fact is that, although the accuracy of DCT features is higher than DWT, its sensitivity is considerably lower. In fact, the highest sensitivity value for textural features (DWT), does not get near to the lowest sensitivity value for the proposed features (set C). This result can be partly explained due to the fact that discarding some scales of the DWT and DCT also means discarding some texture information. In any way histogram measures obtained directly from *a* and *b* color-opponent dimensions of the CIELab color space are superior and avoid heavy computations required by DWT and DCT which speeds up the algorithm. Consequently, in the next subsection only the *a* and *b* color-opponent features will be considered.

6.2.4. Analysis of the segmentation

To analyze the performance of the segmentation module, and because no manual segmentation of the tumors were available, results with and without segmentation will be compared.

In this work the used features were the covariance coefficients between the statistical measures of different regions for each color channel, as described in Section 4.2.1. Therefore, the number of features for each set is six, as can be seen in Figure 39 where R means region (covariances computed between different regions are represented with a dot).

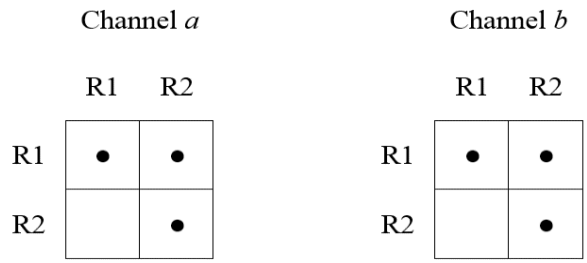


Figure 39. Covariance matrix between both tumoral and normal regions

The use of covariance coefficients, in our view, will improve the features performance, because not just the characteristics of each region (*a* and *b* color channels) will be analyzed, but also the way they vary together between regions which really codes statistical dependence between both regions. Figure 5 shows that these distributions (*a* and *b* color channels) are approximately Gaussian for both normal and tumoral tissue, which means that second order statistics can be enough to characterize them. Second order statistics for independent random variables are just mean and variance. For each region it is not expected that relevant covariance values between color channels exist, since they are inherently independent in the current color space. This means that for example if the light intensity is changed only the L component varies. However, when looking at different regions we expect some degree of statistical dependence between the same random variable (color channel) since the base tissue and subject are the same. This statistical dependence measures perhaps the evolution degree in terms of color of the tumoral tissue from normal tissue. This degree of statistical dependence is partially encoded by covariance measures and characterize unmistakably tumor characteristics not observed in the case where covariances are not computed.

Table 11 and Table 12 present the results with the full set of features for both regions and with the covariance coefficients between regions, respectively.

Table 11. Proposed features taken from two regions using SVM.

Set	Sensitivity	Specificity	Accuracy	AUC
A	88.5	96.8	94.8	95.2
B	86.1	94.2	92.3	91.3
C	67.2	95.9	89.1	86.7
D	85.0	96.2	93.8	91.6

Table 12. Covariance coefficients from Proposed features taken from two regions using SVM.

Set	Sensitivity	Specificity	Accuracy	AUC
A	93.9	96.5	95.9	95.5
B	89.8	91.7	91.2	91.1
C	87.3	90.1	89.4	86.7
D	93.0	95.7	95.1	94.5

Analyzing the results from Table 11 and Table 12, the first conclusion is that in both cases the accuracy values show some improvement when compared to the results without segmentation (Table 9), results consistent with [169].

Although there was an improvement regarding accuracy, it is noticeable a substantial decrease of sensitivity when Table 9 and Table 11 are compared. This was not expected at all and seems to point out that the features from the regions cannot differentiate easily tumoral images from normal images. Because sensitivity is the measure of identifying correctly tumoral frames, low sensitivity values are worst for the overall performance, even if the accuracy is high.

The results are consistent with previous results when we compare the different sets of features, where set A shows the best sensitivity and specificity values (in both cases). Remaining sets present similar behavior as before, with the second best result for set D with a difference of less than 1% in the three measures (Sen, Spe and Acc) when using covariance coefficients. When the full set of region features are used, differences in sensitivity from the best (set A) to the second best (set B) are bigger, being of more than 2% (Table 9).

Sets B and C are again the sets with the worst results. While the sensitivity of set B is almost the same as set D when using the full set of region features, this does not happen with the covariance coefficients, where this difference is of more than 3%, while the differences in specificity are of 2% and 4% respectively. When looking at set C, an extremely low value in sensitivity (below 70%) is obtained, which did not happen in any other experience in this work.

These results show that using covariance coefficients from both regions after segmentation with EM helps to discriminate tumoral tissue in WCE frames. The main purpose to mix coefficients from both regions is to indirectly perform a background normalization procedure. Light variations, both from different devices or in the same device, as well as color variations among subjects can be partially mitigated since they appear in both regions. Both sensitivity and specificity are higher than the ones without segmentation, and in the best case (Set A), the accuracy is increased in more than 1%.

6.2.5. Analysis of the ensemble classifier

In this subsection, the proposed ensemble classifier will be analyzed with different training schemes by differentiating the proposed and the current state-of-the-art approaches of mixture of experts:

1. The training of the gating network is done at the same time as the ensemble elements which is the state-of-the-art mixture of experts. In this approach the entire training dataset is applied to all experts and to the gating network which goal is to mix the experts output (Table 13).
2. Firstly, the training of the ensemble elements is done on the subsets, and only afterwards the gating is trained (Table 14).
3. An online training, where during the test, the ensemble is retrained (Table 15).

In these tables, the results were computed after segmentation of each frame, with the covariance coefficients of the proposed features shown in Table 8 of each region. We also included an experiment with all the 5 statistical measures in a single set (X). Regarding Table 15, it shows the results obtained when incremental adaptation is used for all the incoming test data.

Table 13. Covariance coefficients from Proposed features taken from two regions using Ensemble Classifier with Training scheme 1.

Set	Sensitivity	Specificity	Accuracy	AUC
A	94.8	96.8	96.1	95.3
B	87.2	90.1	89.4	92.1
C	88.2	93.5	92.2	90.2
D	90.8	95.2	94.2	93.5

Table 14. Covariance coefficients from Proposed features taken from two regions using Ensemble Classifier with training scheme 2.

Set	Sensitivity	Specificity	Accuracy	AUC
A	95.1	97.6	97.0	95.8
B	91.6	92.4	92.2	93.2
C	88.7	94.2	92.9	91.6
D	93.2	96.1	95.4	93.2

Table 15. Covariance coefficients from Proposed features taken from two regions using Ensemble Classifier with incremental adaptation for all the test frames (Training scheme 3).

Set	Sensitivity	Specificity	Accuracy	AUC
A	95.6	98.3	97.7	96.4
B	92.1	93.5	92.6	93.8
C	90.2	95.0	93.5	92.3
D	94.1	96.6	95.9	93.7

After concluding that both the proposed features and the segmentation can lead to better results than the textural state-of-the-art features and the entire frame, respectively, in this subsection we will analyze the performance of the ensemble classifier.

By comparing Table 12 (covariance coefficients with an SVM) and Table 13 (covariance coefficients with an ensemble classifier with a common training scheme), we can see that every feature set presented improvements. It is not a surprise that the best results were with set A (when comparing the original four sets), as in every case before. When putting together all the five measures, the results improved since more information was added to the system.

When looking at Table 14, where a different scheme of training is proposed, the results show some improvement. All the sets shown an improvement in the accuracy and sensitivity values, and when looking at AUC, only set C shown a decrease with the different training scheme. This was the expected outcome of this experiment, since high data diversity allows the stabilization of the contribution of each expert along all the dataset. This contribution can have an increase of computation efficiency and system effectiveness if the experts are trained separately from the gating network. In fact, if experts and gating network are trained together, more complex optimization functions are needed, which usually requires approximations that would negatively influence the algorithms convergence.

As can be seen in Table 15 and as was expected, incremental training outperforms static modeling. But this only happens if the diversity is preserved by only adapting the required ensemble element. When all the ensemble elements are retrained, as in regular stack of classifiers, the performance degrades in more than 3% in accuracy, which may happen due to a decrease of diversity.

Overall, the online training leads to an improvement of the results, which shows the advantage of tuning the classifier to the environmental (subject and device related) conditions.

6.3. Conclusion

This work proposes a new approach for tumoral tissue identification in the small bowel based on the RoI concept, with the extraction of covariance coefficients of histogram measures obtained from *a* and *b* color-opponent dimensions of the CIELab color space.

Experimental results show that tumoral tissue can be separated from normal tissue by a kind of lighting normalization accomplished by discarding the L component from Lab color space, which constitutes another novelty presented by this approach. Consequently, tumoral tissue can be assessed by using state-of-the-art image segmentation algorithms, in our case the Bayes approach by using a bi-dimensional (*a* and *b* color-opponent channels) GMM of two components (normal and lesion tissue).Normal frames present inherently normal tissue in both regions (segmentation step forces to divide each image into two regions), which results in poorly separated cases where the convergence of the EM algorithm can be unacceptably slow. In this regard this paper proposes an enhanced formulation of the Anderson acceleration algorithm that outperforms the baseline in more than 10%. Another contribution of this paper is at the classification module

which is ensemble-based, where data partitioning and ensemble structure assure no significant loss in diversity when incremental training is used. Experimental results show the superiority of the proposed algorithm when compared to state-of-the-art procedures in the three main dimensions; feature extraction, tissue separation (automatically performed by segmentation) effectiveness and classification. Improvements of 1.7% and 1.2% were obtained by the proposed feature set and classification module respectively both in accuracy. Additionally, the proposed algorithm requires a significant decrease in computational load when compared with algorithms that require wavelet, curvelet or co-occurrence computations.

Future developments of the algorithm include enhancement of the segmentation module since some of the system failures are originated by segmentation inaccuracies. For that, a manually segmented database must be collected so segmentation metrics can be computed. Best classification schemes must be considered, specifically those based on Deep Learning. The authors are also looking to make clinical tests with this algorithm, applying it to whole exams and comparing the results with clinical reports of WCE.

Chapter 7. Multiple Pathology Detection

This chapter describes the results obtained for a scenario of multi-pathology detection. Most current systems for automatic diagnosis considers the detection of a unique and previously known pathology. Considering specifically the diagnosis of lesions in the small bowel using endoscopic capsule images, very few consider the possible existence of more than one pathology and when they do, they consider only two of them. Such systems do not fully satisfy the medical community, that in fact needs a system that detects any pathology and eventually more than one, when they coexist. In addition, besides the diagnostic capability of these systems, localizing the lesions in the image has been of great interest to the medical community, mainly for training medical personnel purposes. So, nowadays, the inclusion of the lesion location in automatic diagnostic systems is practically mandatory. This chapter was strongly based on [8], which is waiting acceptance at the time of the submission of this thesis.

7.1. Overall Methodology

The methodology used in this regard is the one presented in Section 4.4.3.4. (MI R-CNN) and 4.4.3.5. (training of the model using the second momentum).

The MI R-CNN methodology was proposed to overcome problems with the masks produced with Mask R-CNN that neither PANet nor MS R-CNN could solve. In this way, the proposed approach

uses the information from the RoI alignment subnet in a more efficient way, combining information from different levels of the mask subnet. It is argued that enough information for producing better masks exists, however the mask subnet structure must be improved in order to make a better use of the existing information.

Also, a new methodology was used to train the networks, based on the momentum. Since the momentum is broadly used in the training of CNNs, the use of higher-order momentums was perceived as being a good choice to improve not just the time and resources consumptions when training CNNs, but also reaching an improvement in the performance of the networks.

7.2. Results

Since the proposed methodology is based on both PANet and Mask R-CNN, these two methods were used as baseline for results comparison. Two separate analysis were made. The first one relates to the segmentation analysis, comparing performance values using the different methodologies. A second analysis was made, and related to the classification values, where it is possible to differentiate results not just by methodology, but also by the nature of the lesion.

Regarding the chosen performance metrics, the segmentation analysis was evaluated in terms of mean Average Precision (mAP), which is a metric often used in segmentation-based applications, and its two variants (mAP_{50} and mAP_{75}), which returns the values of the precision considering different values of Intersection over Union (IoU).

$$IoU = \frac{A_{overlap}}{A_{union}} \tag{47}$$

The F1-score metric was also included in this analysis, which uses both precision and recall for its computation.

$$Precision = \frac{TP}{TP + FP} \tag{48}$$

$$Recall = \frac{TP}{TP + FN} \tag{49}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{50}$$

7.2.1. Dataset

The effectiveness of the proposed approach was evaluated in the public database KID dataset 2 [172], [173]. This contains WCE images obtained from the whole GI tract using different exams, all taken with MiroCam® (IntroMedic Co., Seoul, Korea) capsules. These images have a resolution of 360×360 pixels and all were manually annotated and scrutinized by an international scientific committee [172]. These include 303 images of vascular anomalies (small bowel angioectasias, lymphangiectasias, and blood in the lumen), 44 images of polypoid anomalies (lymphoid nodular hyperplasia, lymphoma, Peutz-Jeghers polyps), 227 images of inflammatory anomalies (ulcers, aphthae, mucosal breaks with surrounding erythema, cobblestone mucosa, luminal stenoses and/or fibrotic strictures, and mucosal/villous oedema), and 1,778 normal images obtained from the esophagus, the stomach, the small bowel and the colon.

The vascular anomalies class was divided in two (angioectasias and bleeding), since these two lesions have too many differences in color, size and texture. Also, it is important to refer that for the physicians, the treatment and follow-up of these two lesions are extremely different. In this way, two physicians of Hospital of Braga reviewed the different images and classified them into angioectasia or bleeding separately, and in the case that a classification was different between them, a consensus was reached afterwards.

So, the final dataset used in this work was composed of four different lesion classes: angioectasias, bleeding, polyps and inflammatory lesions; and a class of normal images. Examples of the different lesions present in the dataset can be seen in Figure 40 and the final number of images by lesion are present in Table 16. All frames were used in a proportion of 70% for training and 30% for testing. Training and testing set were randomly selected from the entire dataset and the validation dataset was randomly selected from the training dataset being 15% of the entire dataset.

Table 16. Contents of the used dataset.

Lesion	Number of images
Angioectasia	248
Bleeding	55
Polyps	44
Inflammatory	227
Normal	1,778

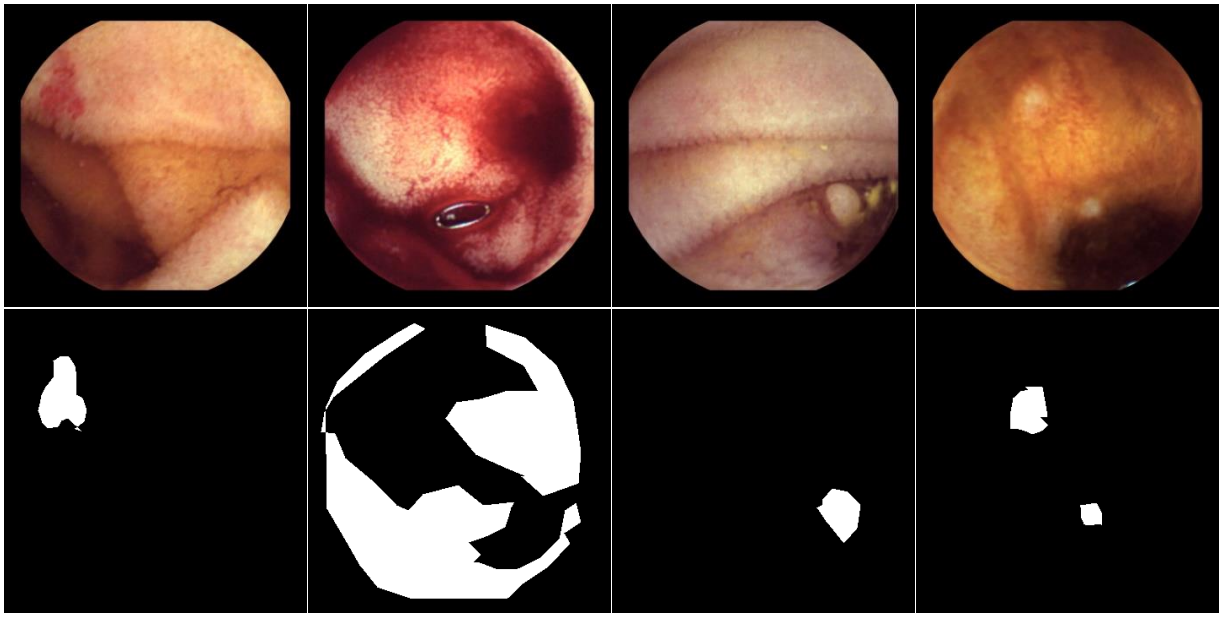


Figure 40. Examples of lesions retrieved from KID Dataset 2. In the top the images and in the bottom the annotated masks. From the left to the right, an example of angioectasia, bleeding, polyp and inflammatory lesion.

7.2.2. Main Results

The considered baseline systems were the Mask-RCNN [159] and PANet [162]. All the results were put together in Table 17, with the presentation of the different metrics (mAP, mAP_{50}, mAP_{75} and F1-score) for the baseline systems, the MI R-CNN and the proposed methodology when using the 2nd momentum for training.

Table 17. Comparison of the results obtained with the unseen test set among Mask R-CNN, PANet, the proposed approach with regular SGD and with the 2nd order momentum optimizer in terms of mAP, mAP_{50}, mAP_{75} and F1-Score.

Method	Backbone	mAP	mAP_{50}	mAP_{75}	F1-score
Mask R-CNN [159]	ResNet-101 + FPN	33.10	56.51	34.71	49.23
PANet [162]	ResNet-101 + FPN	34.66	56.56	36.79	53.66
MI R-CNN	ResNet-101 + FPN	35.75	57.43	38.34	56.83
MI R-CNN + 2nd momentum	ResNet-101 + FPN	**40.35**	**59.42**	**43.01**	**60.07**

Regarding mask refinements, it is important to note that although outperforming the conventional Mask-RCNN in the COCO dataset, the very promising approaches proposed in [160], [161] both underperform the PANet for the KID dataset, therefore for clarity purposes were not shown in Table II. Results show that MI-RCNN with the regular SGD with momentum outperforms the baseline Mask R-CNN and PANet models, with improvements of almost 1% in mAP and mAP_{50}, 4% in mAP_{75} and almost 7% in F1-score. This shows the efficiency of propagating and mixing lower layer information with high level information in the mask subnet. By using the second momentum technique for training purposes, which is the second proposal of this paper and constitutes a novelty for a network of the size of the Mask-RCNN, an extra improvement was achieved. In this case, values of almost 5% in mAP, 2% in mAP_{50}, 5% in mAP_{75} and 3% in F1-score were obtained, which is consistent with the results presented in [163]. Globally, both proposals improve the results of the Mask-RCNN in the KID dataset, which is very significant. Although expecting a significant difference between mask scoring and classification results, the results in this case when looking at the masks shown a strong similarity between classification score and mask alignment, making the system more robust to changes.

When looking at the resultant masks (Figure 41), it is visible that sometimes more than one lesion can be found (first and third example), but in all of these cases, the contour that better filled the annotated mask is the one with the higher probability returned by the model. It is important to note that in these cases, the region with the higher probability was considered to compute the metrics presented in Table 17. It is also possible to understand that the examples with the smaller lesions are the ones where the resultant segmentation have the worst concordance when comparing to the mask, but still better than the state-of-the-art.

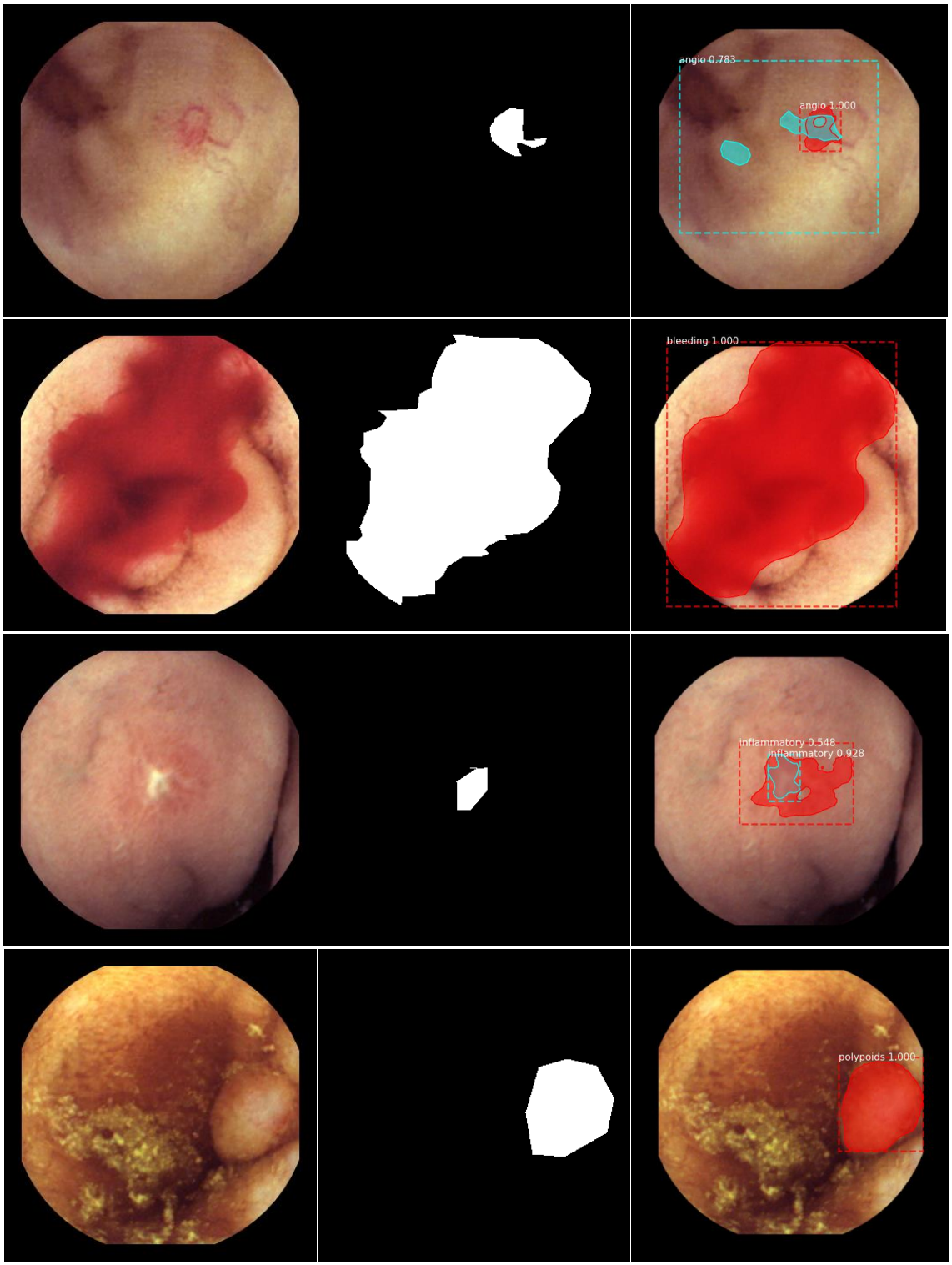


Figure 41. Results of the segmentation using the proposed approach, in examples of the four different considered lesions. From the top to the bottom, an example of angioectasia, bleeding, polyp and inflammatory lesion. From the left to the right, the original image, the annotated mask and the predicted mask with the class probability.

The instance segmentation approach is a fundamental piece in automatic diagnosis for multi-pathology applications, however its use has been very limited which makes it difficult for

performance comparisons with alternative systems as they do not exist. This is the case for multi-pathology detection and lesion segmentation by using WCE videos. In this regard the only possible comparison would be with [173], which is the most advanced known system that uses the full KID Dataset 2. However, as this work does not provide segmentation related metrics, only classification performances comparisons are possible. It is also important to notice that in [173], although using the same dataset, the classification is done as normal/abnormal, and does not classify each lesion individually. Thus, it is not correct to directly compare our multi pathological system with its binary classification system. Table 18 shows the advantages of the instance segmentation approach. In fact, including both classification and segmentation modules in the same system, information sharing is promoted by different subsystems improving their joint use.

Looking at Table 18 it is possible to conclude that the proposed MI-RCNN led to good performances in lesion detection, specially in the case of inflammatory, polypoids and angioectasia lesions. It is noticeable a very promising result with angioectasia detection, which was not the most expected outcome due to the inclusion of other vascular-related lesions in the dataset (bleeding). On the other hand, the results of bleeding detection were substantially lower. By looking at the segmentation we can conclude that some of these lesions were mixed up with shadows and natural hollow spaces of the small bowel. It is also important to note that this was one of the less representative type of lesion in the dataset, which can damage the performance of the classifier. The polypoids detection always achieved the best results when MI-CNN was applied, but a lower precision value was reached when applying the 2nd momentum, which was not expected. But looking at the overall performance, 2nd momentum improved the results when looking at these lesions. Finally, when looking at inflammatory lesions, it is noticeable a significant difference between the precision and the recall values, but the proposed approach reached the best performance of the test.

Overall, when looking at the accuracy of the system, the MI-CNN with the 2nd momentum reached the higher value, and it is a promising result, since no other previous work applied multi-pathology classification and segmentation in the KID dataset. Looking at [173], the only work that used the whole KID dataset 2 for classification purposes, they have reached an accuracy of 77.5%. Although higher than our accuracy of 69.95%, it is difficult to compare both works since their classification was using only a binary classification task of normal/abnormal.

Table 18. Comparison of the classification performances obtained with the unseen test set among Mask R-CNN, PANet, the proposed approach with regular SGD and with the 2nd order momentum optimizer in terms of Recall, Precision, F1-score and Accuracy.

Method	Class	Recall	Precision	F1-score	Accuracy
Mask R-CNN [159]	Inflammatory	45.45	87.5	59.83	62.69
	Polypoids	71.43	90.91	80.00	
	Bleeding	50.00	**100.00**	66.67	
	Angioectasia	79.76	95.71	87.01	
	Weighted Average	62.69	92.49	73.76	
PANet [162]	Inflammatory	45.45	89.74	60.34	63.21
	Polypoids	78.57	84.62	81.48	
	Bleeding	61.11	84.62	70.97	
	Angioectasia	77.38	95.59	85.53	
	Weighted Average	63.21	91.44	73.83	
Proposed MI-RCNN	Inflammatory	51.95	90.91	66.12	66.32
	Polypoids	**78.57**	73.33	75.86	
	Bleeding	**66.67**	85.71	**75.00**	
	Angioectasia	77.38	94.20	84.97	
	Weighted Average	66.32	90.58	75.86	
Proposed MI-RCNN + 2nd momentum	Inflammatory	**55.84**	**93.48**	**69.92**	**69.95**
	Polypoids	71.43	**100.00**	**83.33**	
	Bleeding	61.11	73.33	66.67	
	Angioectasia	**84.52**	**97.26**	**90.45**	
	Weighted Average	**69.95**	**93.72**	**79.52**	

7.3. Conclusions

Current clinical practice of the gastrointestinal tract requires multi-pathology detection given the amount of different pathologies that can be found. Lesion localization modules provided by modern automatic diagnosis systems have been highly appreciated by the medical community, since the specific region of the lesion is shown, improving the physician confidence in the system. Therefore, a useful system must have a classification module and a segmentation module that can

complement each other if they share components. Current instance segmentation systems have these characteristics, while Mask-RCNN being one of the most used systems of this type. One of the characteristics that can be improved in Mask-RCNN is the quality of the predicted mask. MS-RCNN and BMask-RCNN are two methods that improve the predicted mask quality in the COCO dataset, however both underperform PANet in our case. PANet adds a branch in the mask sub-net containing tiny fully connected layers that can capture different views of each proposal increasing information diversity hence producing masks of better quality.

This paper proposed MI-RCNN, an efficient method to improve the quality of the predicted mask that outperforms PANet. The method is based on the well-established principle that forwarding low-layer information and mixing this information with high-level information makes a better use of the existing information. In this regard, propagation of low layer information from all sub-net mask levels to the C4_fc layer is proposed. Also, the use of the 2nd moment for training the network instead of a simple momentum was proposed, which is an innovative contribution from this work.

Experimental results show that the proposed methods significantly improve the evaluation metrics, with an increase of 3% in F1-score. By training the proposed model with the innovative method based on the second momentum an extra improvement of more than 3% was achieved over the PANet model. The classification results also followed the same behavior, with accuracies for the proposed MI-RCNN and MI-RCNN with 2nd momentum 3% and 7% higher than the PANet, respectively, for the majority of the analyzed lesions.

Although it is a first work in a scenario of multi-pathology detection in WCE images, the achieved results are quite promising. The good results not only in the classification task, but also on the segmentation task, could lead to the conclusion that in fact, these two modules should be always complementary to each other. As future work, we would like to improve the masks predictions to achieve better results, increase the number of pathologies by using other datasets and validate in a clinical setting the whole system.

Chapter 8. Clinical Validation

This chapter describes the work done as a part of a Medical Student Project that began in 2017 as a Project of the 4th Year and then evolved to her Master's Degree Dissertation. The work was done in the School of Medicine of University of Minho, Hospital of Braga and School of Engineering. The main supervisor of the work was Dalila Costa, M.D. and I was the co-supervisor. This work was published in [9] and presented in the National Congress of Gastroenterology of 2019 where it won the National Gastroenterology Prize.

Only the angioectasias detector described in Chapter 5 was used, because it was the only software finished at the time for testing.

8.1. Protocol

To assess the usability of the algorithm in clinical practice, WCE recordings with SB angioectasias consecutively performed (2017-2018) in the Department of Gastroenterology of Hospital of Braga were evaluated (which consisted in a total of 41 WCE videos). Videos with blood

and poor preparation were excluded. WCE procedures were performed according to a standardized protocol: clear liquid diet the day prior to VCE plus a 12-hour fasting protocol. The recording device was set in the morning and removed in the late afternoon. Patients were allowed to drink fluids after 2 hours and to have a light meal after 4 hours.

All WCE recordings were firstly interpreted by Physician 1 (an expert capsule endoscopist with an experience around 500 capsules procedures), who proposed the clinical diagnosis of angioectasias and also registered the anatomic landmarks, the clinically significant lesions, time required for viewing, and quality of SB cleansing. Only lesions in the SB were considered (those found in between the pylorus and the ileocecal valve as defined by the first duodenal and first colonic video frames, respectively).

Posteriorly, two gastroenterology trainees with different WCE reading experience (50 and 10 capsules procedures; Physician 2 and Physician 3, respectively) were asked to review the VCE recordings and to register the data described in the previous paragraph for Physician 1. The SingleView mode of VCE was defined for all physicians. Each physician reviewed each VCE only once and without knowing the results from either one of the other physicians and before applying the algorithm.

WCE videos were downloaded from the original software. The algorithm was applied separately from the vendors' software and provided a smaller video consisting of the grouping of frames selected. All images tagged by the algorithm and physicians were reviewed by an independent capsule endoscopist (experience around 150 CE) that confirmed if the lesion selected was an angioectasia and evaluated the concordance between them.

Afterwards some metrics were computed, specifically physicians' performance, regarding Global Detection Rate (total number of angioectasias detected), Detection Rate per capsule (number of angioectasias tagged in each WCE), Diagnostic yield (number of WCE with clinical diagnosis of angioectasias), True Positive Rate (TPR, number of lesions correctly identified as angioectasia) and viewing time.

At last, an analysis was also made to assess the performance of each physician together with the results of the algorithm. The purpose was to infer what would be the result of using the algorithm in a clinical practice as a complement to physicians' readings of WCEs.

The experimental protocol was submitted and approved by the Sub-commission of Life and Health Sciences, University of Minho, and Ethics Committee for Health of Braga Hospital.

8.1.1. Statistical analysis

Statistical analysis was performed using Statistical Package for the Social Sciences (SPSS Inc®, Chicago, Illinois, USA version 24.0 for Microsoft® Windows). For the descriptive data, absolute frequencies (n) and relative frequencies (%) were used for qualitative variables; and medians and interquartile range (IQR) were used for quantitative variables. Mean and standard deviations were also presented for quantitative variables, when required to facilitate data interpretation. Normal distribution was assessed by Skewness & Kurtosis tests. Since this assumption was not observed, non-parametric statistics were used throughout the analysis.

Mann-Whitney (U), Wilcoxon signed-rank (Z) and Friedman test were performed to compare the distribution of continuous variables. Chi-Square ($\chi 2$), Fisher's and Cochran's Q test ($\chi 2$) were assessed to test the association between categorical dichotomic variables Results were considered statistically significant for a p value below 0.05.

8.2. Results

The section of the results will be explored in different phases. First, a overall performance of the software using the chosen WCE exams was performed, where the rate of angioectasia detection of the software was attained. Afterwards, the results from physicians analysis was done, since it was the main objective of this chapter.

8.2.1. Overall performance of the software

In the evaluation of the full-length 54 WCE recordings, the algorithm revealed a detection rate of 77.70% (identified 115 from a total of 148 angioectasias). The lower performance significantly resulted from the failure of the algorithm to identify punctiform angioectasias ($p<0.001$).

No significant differences were observed in the algorithm performance between MiroCam® (74.2%) and PillCam® (78.6%) ($p = 0{,}598$). Regarding the number of frames (and subsequently the size of the video performed by the software), the algorithm significantly selected more frames

in WCEs from MiroCam® (M=886.06 v 372.45; p = 0,028), suggesting a significant lower specificity, but maybe partially explained due to the higher size of videos from MiroCam® WCEs.

To understand the differences between the images from the different capsules and to understand what these can influence in the lesion detection algorithm performance, an experiment was made. A group of normal SB images (100 images) were randomly selected from each WCE type and histograms with pixel intensities of both *a* and *b* color channels are shown in Figure 42. Both *a* ($p<0.00001$) and *b* ($p<0.00001$) channels of Mirocam® frames significantly showed a higher intensity, with images displaying a more yellowish and reddish coloration, resulting in more false positives.

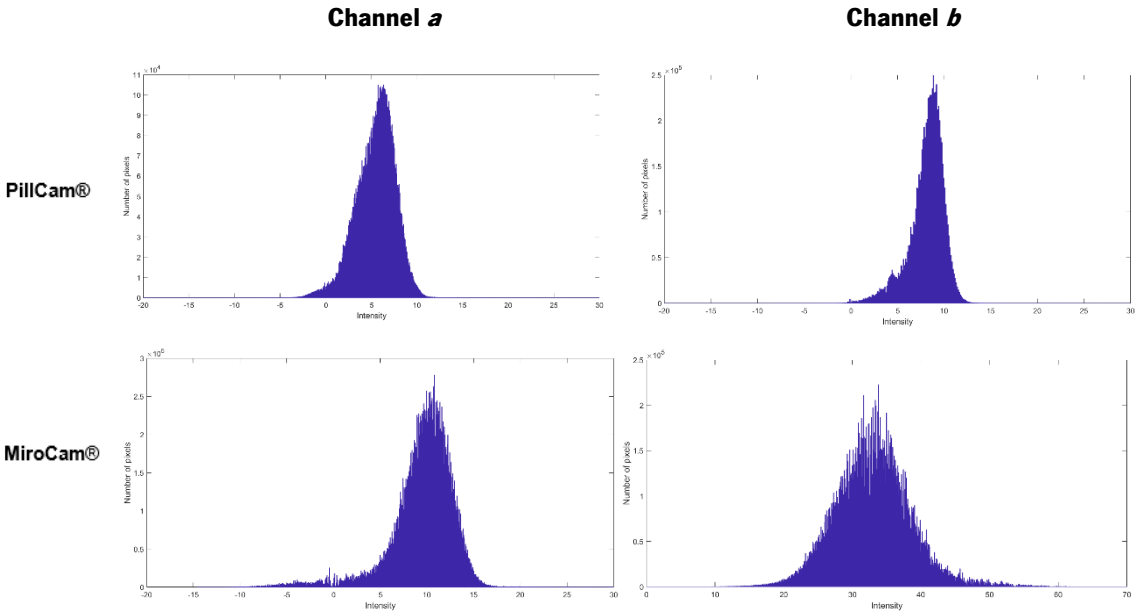


Figure 42. Histograms of pixels from randomly selected 100 frames, from the two WCE's models and from the two-color channels (*a* and *b*).

In Figure 43 and Figure 44 some frames from the different WCE brands are shown (PillCam® and MiroCam®, respectively). Although only a sample of frames is shown, it is clear that the coloration differs from the different WCE brands, specially regarding the increase of yellow coloration in MiroCam® frames. Regarding the false negative results, both are explained with the presence of bubbles in the images and the reduced size of the lesion. Otherwise, the false positives in both figures can be explained with the presence of blood vessels in the region, which have a similar color than the lesions.

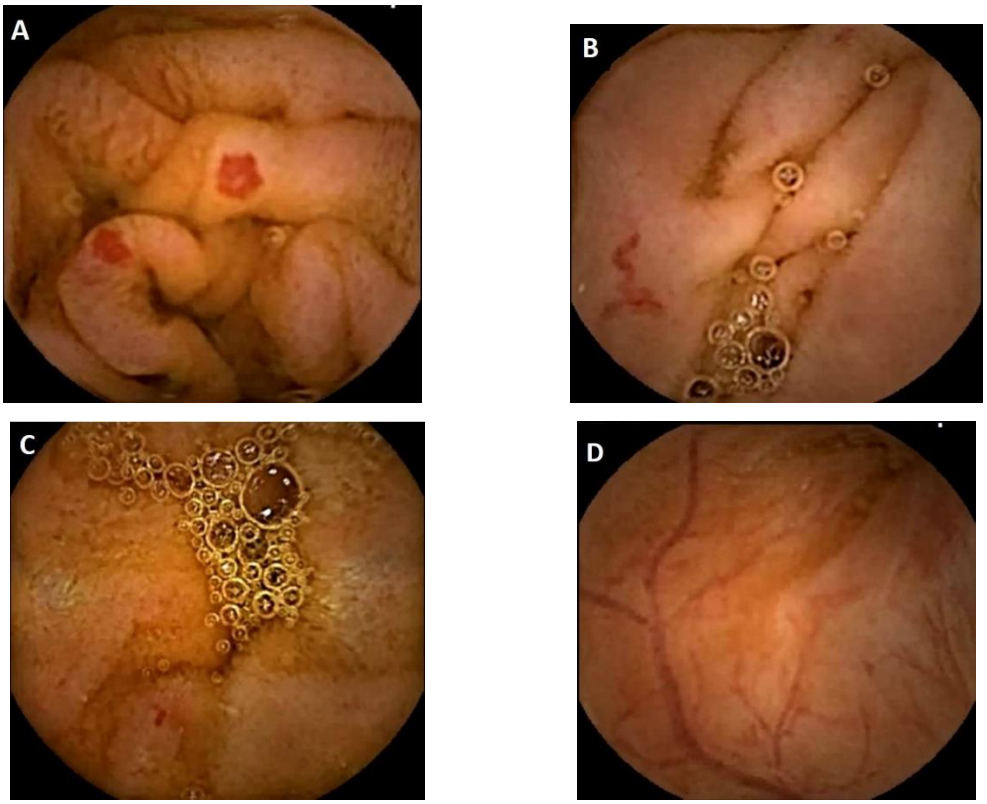


Figure 43. Examples of frames from PillCam® software selected by the algorithm: (a) and (b) are true positive, (c) is a false negative and (d) is a false positive.

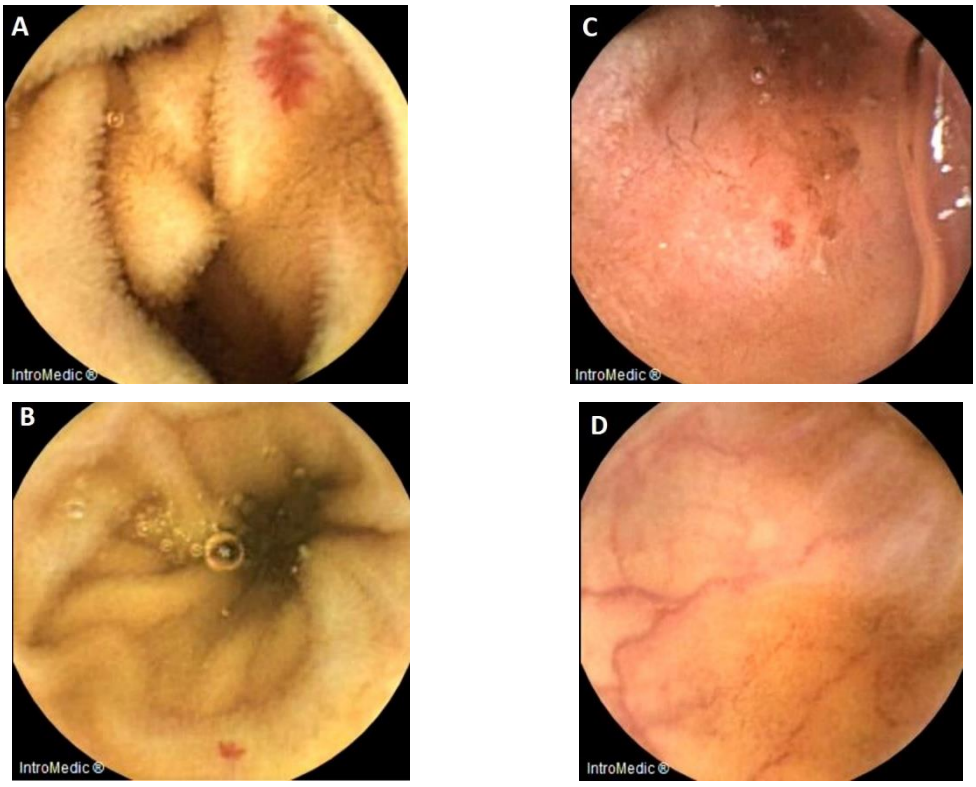


Figure 44. Examples of frames from MiroCam® software selected by the algorithm: (a) and (b) are true positive, (c) is a false negative and (d) is a false positive.

8.2.2. Overall performance of the Physicians

To overcome the limitation inherent to most studies that examine CAD tools, the algorithm performance was evaluated in full-length WCE recordings. As expected, angioectasias detection rate diminished but a satisfactory performance was maintained (77.70%). This result was still superior to SBI, which detain a sensitivity of 55% to detect potentially bleeding lesions. Considering the 38 WCE recordings selected, the algorithm revealed a diagnostic yield of 94.70% (n=36). The 2 WCE examinations that the algorithm failed to identify any lesion, displayed an isolated punctiform angioectasia. Interestingly, the software mostly failed to identify punctiform angioectasias (reduced area/volume). Nevertheless, it identified 3 angioectasias that were not identified by none of the 3 selected physicians. Considering the positive correlation between severity and lesion extension [6], [174], the algorithm sustained a great detection rate of lesions with higher probability of bleeding.

Concerning the Global Detection Rate (Table 19), no significant differences were observed between Physician 1 and Physician 2 ($p=0.728$), but both were significantly superior than Physician 3 ($p<0.001$) After addition of the algorithm analysis, each physician significantly improved the total number of angioectasias detected ($p<0.001$) (Table 20) and differences between physicians' performance were no longer noted ($p=0.108$). Similar results were obtained in Detection Rate per capsule. Physician 1 and Physician 2 did not significantly differ ($p=0.064$), though both revealed greater detection rate than Physician 3 ($p<0.001$ and $p=0.029$, respectively) (Table 19). All physicians revealed significantly better results following algorithm application (Table 20) and differences between them were diluted ($p=0.108$).

In regard to diagnostic yield, only Physician 1 detained significant superiority against Physician 3 ($p=0.008$) (Table 19). Both Physician 2 and Physician 3 improved their diagnostic yield, though without significance ($p=0.125$ and $p=0.063$, respectively) (Table 20). Nevertheless, differences between physicians' performance were no longer detected following algorithm inclusion ($p=0.097$).

At last, Physician 1 revealed significantly higher TPR than Physician 2 ($p<0,001$) and Physician 3 ($p<0.001$). Physician 2 also displayed a higher TPR value when compared to Physician 3 ($p=0.003$) (Table 20). Similar results were observed regarding WCE viewing time. Physician 1 was faster than the other physicians ($p<0.001$). Physician 3 was also significantly slower than

Physician 2 (p<0.001), but no significant differences were revealed between the time spent in WCE viewing.

Table 19. Physicians performance assessment.

	Physician 1	Physician 2	Physician 3	Test Statistics
Global DR				
% (n)	81.70 (94)	79.10 (91)	53.90 (62)	X^2=30.23 p<0.001*
DR per capsule (%)				
Mean (+/- SD)	90.06 (15.71)	76.70 (34.77)	58.63 (38.28)	X^2 = 20.80
Median (IQR)	100 (25.89)	100 (43.25)	53.57 (66.67)	p<0.001**
Diagnostic yield				
% (n)	100 (38)	86.8 (33)	78.9 (30)	X^2 = 9.80 p=0.007*
TPR				
Mean (+/- SD)	—[1]	100 (42.50)	51.78 (35.39)	X^2 = 39.08
Median (IQR)	—[1]	74.85 (37.23)	50 (56.43)	p<0.001**
Time (min)				
Mean (+/- SD)	25.26 (5.11)	40.76 (22.40)	44.63(25.30)	X^2 = 41.45
Median (IQR)	25 (8)	36.50 (20)	40 (21)	p<0.001**

DR: Detection Rate; TPR: True Positive Rate; SD: Standard Deviation; IQR: Interquartile Range.

—1 - This value is a constant for this physician, 100%.

—2 - This value is a constant for this physician, 20.

* Cochran's Q test (McNemar tests with Bonferroni correction are displayed in the text).

** Friedman test (Wilcoxon tests with Bonferroni are presented in the text).

Table 20. Physicians performance with and without the software.

	Physician 1	+A	Physician 2	+A	Physician 3	+A
DR (angioectasias)						
% (n)	81.70 (94)	92.20 (106)	79.10 (91)	95.70 (110)	53.90 (62)	89.6 (103)
Test Statistics	p<0.001*		p<0.001*		p<0.001*	
DR per capsule (%)						
Mean (+/-SD)	90.06 (15.71)	95.72 (10.78)	76.70 (34.77)	92.31 (22.66)	58.63 (38.28)	88.19 (25.07)
Test Statistics	p=0.012**		p =0.004**		p<0.001**	
Diagnostic yield						
% (n)	100 (38)	100 (38)	86.8 (33)	97.4 (37)	78.9 (30)	92.1 (35)
Test Statistics	-		p=0.125*		p=0.063*	

*Wilcoxon test;

** Mann Whitney test.

Note: It is impossible to compare Diagnostic yield of Physician 1, because WCEs were selected according to the clinical diagnosis of Angioectasias performed by Physician 1.

A: Algorithm; p : p-value; SD: Standard Deviation.

8.3. Conclusion

This is the first study to assess performances (both from physicians and a software) using different WCE systems. Most available tools to enhance WCE diagnostic yield were developed using Given Imaging®, though other systems have been widely used in the clinical practice. Software performance did not differ between both endoscopic capsules evaluated. Even though specificity was not possible to calculate, our data suggests that our algorithm displays a lower specificity in WCE recordings from Mirocam®, since the number of frames selected by the algorithm was significantly higher (at the expense of a heavier video). This occurs as a consequence of Mirocam® frames displaying a more yellowish and reddish coloration, resulting in more false positives. Since the algorithm was trained with PillCam® images, it possibly makes sense to change the segmentation method so that the software could better separate the tissue from the angioectasias amongst the normal tissue. Despite this, the algorithm proved to be able to detect angioectasias in both systems, which allows to infer that following a training with images of both capsules systems, the algorithm will increase its diagnostic performance.

The software revealed an excellent diagnostic yield (94.70%), though interpretation should be carefully done due to the small sample used. Notwithstanding, the data may be comparable or even better than previously discussed tools [167], [175]–[178]. Arieira *et al* revealed an overall diagnostic yield of 42% for Top100, though angioectasia was diagnosed in 86,7% [179]. Carvalho *et al* also showed that FICE 1 displayed a significantly higher diagnostic yield than white light (55% vs 42%), particularly to identify angioectasias [180]. Once more, the 2 WCE examinations that the algorithm failed to identify any lesion displayed an isolated punctiform angioectasia. Hence, training the algorithm not only will improve algorithm overall performance, but also its diagnostic yield.

The main purpose of these computerized tools is to enhance physicians' performance in the clinical practice. Zheng *et al* revealed that the detection rate of angioectasias by physicians is disappointing, since only 69% of these lesions were detected [52]. In our study, the most experienced physician was faster to review CE recordings while revealing greater diagnostic yield and TPR than both trainees, but his Global Detection Rate and Detection Rate per capsule were not superior to the most experienced trainee. Even though the achieved results should not be inferred to the population due to the small sample; our data is comparable to the literature. Experienced capsule endoscopists are not necessarily better interpreters than their less-

experienced counterparts [52], [181]. Therefore, experience should not serve as a pass to review studies more quickly.

Remarkably, the software enhanced physicians' performance in the clinical practice. Following the addition of algorithm analysis, all readers significantly improved their Global Detection Rate and Detection Rate per capsule. Moreover, significant differences between them were no longer observed. Consequently, physicians with different experiences in WCE examinations may achieve similar clinical performance, which may also suggest a role of this software in the learning process of the Gastroenterology trainee.

The main limitation was sample size and the number of physicians enrolled in the study. More than one study independent capsule endoscopist should have been used to achieve a more accurate algorithm and readers' performance data. In the future, we intend to improve this tool in order to overcome the exposed failures. Also, the expected step is to develop an improved version of the software where other lesions can be simultaneously detected.

In conclusion, this new computerized tool displayed an excellent diagnostic yield and improved physicians' performance in clinical practice. However, multicenter and randomized controlled trials are important to validate these findings.

Chapter 9. Conclusions and Future Work

In this chapter, main conclusions of the developed work will be presented, along with some proposals for future work in the area of lesion detection of WCE exams.

9.1. Main Conclusions

The development of automatic or semi-automatic systems with the purpose of helping physicians to correctly diagnose lesions has systematically grown in the past few years. The technology and scientific advances in computation, electronics and mathematics have been one of the reasons for it. It is clear that the arrival of smaller and more powerful processors was one of the main reasons for the flourish of new and innovative methods around Machine Learning, and specially Deep Learning. One other reason is a better link with the physicians, with a higher understanding that technology can indeed improve their practice, reaching better clinical results and leaving some tedious tasks for AI applications and other IT innovative solutions.

With the increase of the prevalence of GI lesions and diseases around the world, there is the need of focus on the research and the increase of investment in this specific area of healthcare.

The WCE device, although having 20 years old, has still a strong room to increase its usage, with predictions of WCE market to reach around 401 million dollars by 2028 (the value for 2020 has not reached 300 million dollars) [182]. Having this into consideration, it is important to notice that this low usage can have various factors, from the (a) high cost of each capsule, (b) need for experience readers within the hospitals, (c) need to analyze the video, which is a tedious task. The inclusion of new features in WCE devices, for example lesion detection modules, can help to increase WCE usage around the world, and possibly reduce the costs associated to it.

This thesis tried to improve the research done in the field of automatic lesion detection systems for WCE exams. The objective was to tackle different challenges in this field, from the detection of specific lesions (like angioectasias or tumors) to the detection of multiple pathologies at the same time. One of the objectives was also to try to apply all or some of the developments in the clinical practice, which was partially achieved with a study of angioectasia detection system applied in the Hospital of Braga.

With this work, it could be concluded that a single learning system detecting only angioectasias could in fact reach similar diagnosis that an expert physician, when looking to a database of several dozens of exams. It was also proven that in cases of lesions with defined borders and small size, segmentation can be important for a correct detection. Alongside this result, it was shown that a new ensemble classifier increased the performance of a tumor detection system, specially using a newly designed online training scheme. A first approach dedicated to multiple pathology detection systems was developed during this thesis, with the use of a new scheme based on Mask R-CNN, which shown that indeed it is possible to achieve promising results when looking at the performance of the proposed methodology.

9.2. Future Work

Regarding future work, there are still a lot of challenges to overcome in this field.

One major challenge is the lack of public databases which are big and heterogenous enough to allow researchers to test their method and compare them in the best possible way. This should be done including images and videos of different WCE systems (as was explained, different WCE brands have different characteristics which leads to different type of images), and lesions with a representative number of images, with GT masks done by more than one physician. A similar database is being developed (CAD-CAP) [183], but at the time of submission of this thesis is not

yet publicly available. The public database with the most number of images and different lesions available is KID [172], which was used throughout this thesis.

Other important area that needs to be addressed is the clinical study of the multi-pathology system that was developed in this thesis. Although the clinical work done during this thesis presented relevant results, it is needed to make a more broaden study, using not just the system that has the ability of detect more than one GI pathology, but also with a protocol where it can be used integrated in the software used by physicians and directly in the WCE exams that are being performed in the Hospital. Of course this procedure would need the collaboration of a large number of gastroenterologists, preferably from different hospitals so a big number of exams could be reached.

References

[1] A. F. Peery *et al.*, "Burden of Gastrointestinal Disease in the United States: 2012 Update," *Gastroenterology*, vol. 143, no. 5, pp. 1179-1187.e3, Nov. 2012.

[2] M. Farthing *et al.*, "Survey of digestive health across Europe: Final report. Part 1: The burden of gastrointestinal diseases and the organisation and delivery of gastroenterology services across Europe," *United Eur. Gastroenterol. J.*, vol. 2, no. 6, pp. 539–543, Dec. 2014.

[3] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, no. 6785, p. 417, 2000.

[4] W. A. Qureshi, "Current and future applications of capsule endoscopy," *Nat. Rev. Drug Discov.*, vol. 3, pp. 447–450, 2004.

[5] P. M. Vieira, C. P. Silva, D. Costa, I. F. Vaz, C. Rolanda, and C. S. Lima, "Automatic Segmentation and Detection of Small Bowel Angioectasias in WCE Images," *Ann. Biomed. Eng.*, vol. 47, no. 6, pp. 1446–1462, Jun. 2019.

[6] P. M. Vieira, B. Goncalves, C. R. Goncalves, and C. S. Lima, "Segmentation of angiodysplasia lesions in WCE images using a MAP approach with Markov Random Fields," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 1184–1187.

[7] P. M. Vieira, N. R. Freitas, J. Valente, I. F. Vaz, C. Rolanda, and C. S. Lima, "Automatic detection of small bowel tumors in wireless capsule endoscopy images using ensemble learning," *Med. Phys.*, p. mp.13709, Nov. 2019.

[8] P. M. Vieira, N. R. Freitas, V. B. Lima, D. Costa, C. Rolanda, and C. S. Lima, *Multi-pathology Detection and Lesion Localization in WCE Videos by Using the Instance Segmentation Approach*. Manuscript Submitted for publication, 2020.

[9] D. Costa *et al.*, "Clinical Performance of New Software to Automatically Detect Angioectasias in Small Bowel Capsule Endoscopy," *GE - Port. J. Gastroenterol.*, pp. 1–10, Oct. 2020.

[10] W. F. Boron and E. L. Boulpaep, *Medical Physiology: A Cellular and Molecular Approach*. Saunders, 2003.

[11] F. C. Brunicardi, T. R. Billiar, D. L. Dunn, J. G. Hunter, and R. E. Pollock, *Schwartz' Principles of Surgery : Self-Assessment and Board Review, Eighth Edition: Self-Assessment and Board Review, Eighth Edition*. Mcgraw-hill, 2006.

[12] H. Gray, *Anatomy of the human body*. New York, NY, USA: Philadelphia: Lea & Febiger, 1918.

[13] M. Pennazio *et al.*, "Outcome of patients with obscure gastrointestinal bleeding after capsule endoscopy: Report of 100 consecutive cases," *Gastroenterology*, vol. 126, no. 3, pp. 643–653, Mar. 2004.

[14] R. Berkow and M. H. Beers, *Manual Merck - Saúde para a família*. Oceano Grupo Editorial, S.A., 2003.

[15] M. Keuchel, N. Kurniawan, and P. Baltes, "Small bowel ulcers," *Curr. Opin. Gastroenterol.*, vol. 35, no. 3, pp. 213–222, May 2019.

[16] A. Karargyris and N. Bourbakis, "Detection of Small Bowel Polyps and Ulcers in Wireless Capsule Endoscopy Videos," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2777–2786, Oct. 2011.

[17] U. S. Karnam, C. M. Rosen, and J. B. Raskin, "Small bowel ulcers," *Curr. Treat. Options Gastroenterol.*, vol. 4, no. 1, pp. 15–21, Feb. 2001.

[18] K. J. Lin, L. A. García Rodríguez, and S. Hernández-Díaz, "Systematic review of peptic ulcer disease incidence rates: do studies without validation provide reliable estimates?," *Pharmacoepidemiol. Drug Saf.*, vol. 20, no. 7, pp. 718–728, Jul. 2011.

[19] A. P. Hemingway, "Angiodysplasia: current concepts.," *Postgrad. Med. J.*, vol. 64, no. 750, pp. 259–263, Apr. 1988.

[20] J. Regula, E. Wronska, and J. Pachlewski, "Vascular lesions of the gastrointestinal tract," *Best Pract.*

Res. Clin. Gastroenterol., vol. 22, no. 2, pp. 313–328, Apr. 2008.

[21] G. W. Fan *et al.*, "Angiodysplasia and bleeding in the small intestine treated by balloon-assisted enteroscopy," *J. Dig. Dis.*, vol. 14, no. 3, pp. 113–116, Mar. 2013.

[22] S. S. Sami, S. A. Al-Araji, and K. Ragunath, "Review article: gastrointestinal angiodysplasia - pathogenesis, diagnosis and management," *Aliment. Pharmacol. Ther.*, vol. 39, no. 1, pp. 15–34, Jan. 2014.

[23] T. E. Warkentin, J. C. Moore, S. S. Anand, E. M. Lonn, and D. G. Morgan, "Gastrointestinal bleeding, angiodysplasia, cardiovascular disease, and acquired von Willebrand syndrome," *Transfus. Med. Rev.*, vol. 17, no. 4, pp. 272–286, Oct. 2003.

[24] D. Gunjan, V. Sharma, S. S. Rana, and D. K. Bhasin, "Small bowel bleeding: a comprehensive review," *Gastroenterol. Rep.*, vol. 2, no. 4, pp. 262–275, Nov. 2014.

[25] A. Szold, L. B. Katz, and B. S. Lewis, "Surgical approach to occult gastrointestinal bleeding," *Am. J. Surg.*, vol. 163, no. 1, pp. 90–93, Jan. 1992.

[26] K. Otani *et al.*, "Usefulness of small bowel reexamination in obscure gastrointestinal bleeding patients with negative capsule endoscopy findings: Comparison of repeat capsule endoscopy and double-balloon enteroscopy," *United Eur. Gastroenterol. J.*, vol. 6, no. 6, pp. 879–887, Jul. 2018.

[27] W. Honda *et al.*, "Enteroscopic and radiologic diagnoses, treatment, and prognoses of small-bowel tumors," *Gastrointest. Endosc.*, vol. 76, no. 2, pp. 344–354, Aug. 2012.

[28] R. A. de Latour, S. M. Kilaru, and S. A. Gross, "Management of small bowel polyps: A literature review," *Best Pract. Res. Clin. Gastroenterol.*, vol. 31, no. 4, pp. 401–408, Aug. 2017.

[29] E. Chen and G. Vaccaro, "Small Bowel Adenocarcinoma," *Clin. Colon Rectal Surg.*, vol. 31, no. 05, pp. 267–277, Sep. 2018.

[30] Y.-H. Hsieh and F. Leung, "Increase your adenoma detection rate without using fancy adjunct tools," *Tzu Chi Med. J.*, vol. 30, no. 3, p. 127, 2018.

[31] C. Van de Bruaene, "Small bowel capsule endoscopy: Where are we after almost 15 years of use?," *World J. Gastrointest. Endosc.*, vol. 7, no. 1, p. 13, 2015.

[32] D. Y. Cheung and M.-G. Choi, "Current Advance in Small Bowel Tumors," *Clin. Endosc.*, vol. 44, no. 1, p. 13, 2011.

[33] S. Sripathi, R. Srivastava, K. Rajagopal, and A. Ayachit, "CT features, mimics and atypical presentations of gastrointestinal stromal tumor (GIST)," *Indian J. Radiol. Imaging*, vol. 21, no. 3, p. 176, 2011.

[34] H. Yamamoto *et al.*, "Total enteroscopy with a nonsurgical steerable double-balloon method," *Gastrointest. Endosc.*, vol. 53, no. 2, pp. 216–220, Feb. 2001.

[35] V. S. Kodogiannis, M. Boulougoura, E. Wadge, and J. N. Lygouras, "The usage of soft-computing methodologies in interpreting capsule endoscopy," *Eng. Appl. Artif. Intell.*, vol. 20, no. 4, pp. 539–553, Jun. 2007.

[36] T. Nakamura and A. Terano, "Capsule endoscopy: past, present, and future," *J. Gastroenterol.*, vol. 43, no. 2, pp. 93–99, Feb. 2008.

[37] M. R. Basar, F. Malek, K. M. Juni, M. S. Idris, and M. I. M. Saleh, "Ingestible Wireless Capsule Technology: A Review of Development and Future Indication," *Int. J. Antennas Propag.*, vol. 2012, pp. 1–14, 2012.

[38] P. K. Sa, M. N. Sahoo, M. Murugappan, Y. Wu, and B. Majhi, Eds., *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*, vol. 719. Singapore: Springer Singapore, 2018.

[39] R. Martin, "Can Capsule Endoscopy Technology Keep You Out of Surgery?," *Ignite*, 2018. .

[40] C. Li, B. Zhang, C. Chen, and Y. Li, "OMOM capsule endoscopy in diagnosis of small bowel disease," *J. Zhejiang Univ. Sci. B*, vol. 9, no. 11, pp. 857–862, Nov. 2008.

[41] J. A. Barkin and J. S. Barkin, "Video Capsule Endoscopy," *Gastrointest. Endosc. Clin. N. Am.*, vol. 27, no. 1, pp. 15–27, Jan. 2017.

[42] H. VU *et al.*, "Controlling the Display of Capsule Endoscopy Video for Diagnostic Assistance," *IEICE Trans. Inf. Syst.*, vol. E92-D, no. 3, pp. 512–528, 2009.

[43] S. Liangpunsakul, L. Mays, and D. K. Rex, "Performance of given suspected blood indicator," *Am. J. Gastroenterol.*, vol. 98, no. 12, pp. 2676–2678, Dec. 2003.

[44] S. A. Zanati *et al.*, "Value of the Suspected Blood Indicator in Wireless Capsule Endoscopy," *Gastrointest. Endosc.*, vol. 59, no. 5, p. P167, Apr. 2004.

[45] D. E. Yung, C. Sykes, and A. Koulaouzidis, "The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis," *Expert Rev. Gastroenterol. Hepatol.*, vol. 11, no. 1, pp. 43–51, Jan. 2017.

[46] P.-N. D'Halluin *et al.*, "Does the 'Suspected Blood Indicator' improve the detection of bleeding lesions by capsule endoscopy?," *Gastrointest. Endosc.*, vol. 61, no. 2, pp. 243–249, Feb. 2005.

[47] P. Boal Carvalho *et al.*, "Suspected blood indicator in capsule endoscopy: a valuable tool for gastrointestinal bleeding diagnosis," *Arq. Gastroenterol.*, vol. 54, no. 1, pp. 16–20, Mar. 2017.

[48] Given Imaging, "PillCam Capsule Endoscopy User Manual," Mansfield, USA, 2016.

[49] C. Arieira *et al.*, "Su1259 Is the Software Top 100 a Reliable Tool in Capsule Endoscopy?," *Gastrointest. Endosc.*, vol. 87, no. 6, p. AB301, Jun. 2018.

[50] C. Arieira *et al.*, "Capsule endoscopy: Is the software TOP 100 a reliable tool in suspected small bowel bleeding?," *Dig. Liver Dis.*, Jul. 2019.

[51] U. Günther, S. Daum, M. Zeitz, and C. Bojarski, "Capsule endoscopy: comparison of two different reading modes," *Int. J. Colorectal Dis.*, vol. 27, no. 4, pp. 521–525, Apr. 2012.

[52] Y. Zheng, L. Hawkins, J. Wolff, O. Goloubeva, and E. Goldberg, "Detection of lesions during capsule endoscopy: physician performance is disappointing.," *Am. J. Gastroenterol.*, vol. 107, no. 4, pp. 554–60, Apr. 2012.

[53] A. Novozámský, J. Flusser, I. Tachecí, L. Sulík, J. Bureš, and O. Krejcar, "Automatic blood detection in capsule endoscopy video," *J. Biomed. Opt.*, vol. 21, no. 12, p. 126007, Dec. 2016.

[54] "Capsule Endoscopy Small Bowel Disorders - Fast, Accurate Diagnosis with Capsule Endoscopy," *SynMed*. [Online]. Available: http://www.synmed.co.uk/products_capsule_endoscopy.htm. [Accessed: 31-Oct-2019].

[55] D. K. Iakovidis, S. Tsevas, D. Maroulis, and A. Polydorou, "Unsupervised summarisation of capsule endoscopy video," in *2008 4th International IEEE Conference Intelligent Systems*, 2008, pp. 3-15-3–20.

[56] D. K. Iakovidis, S. Tsevas, and A. Polydorou, "Reduction of capsule endoscopy reading times by unsupervised image mining," *Comput. Med. Imaging Graph.*, vol. 34, no. 6, pp. 471–478, 2010.

[57] M. M. Ben Ismail, O. Bchir, and A. Z. Emam, "Endoscopy video summarization based on unsupervised learning and feature discrimination," in *2013 Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.

[58] Z. Sun, B. Li, R. Zhou, H. Zheng, and M. Q.-H. Meng, "Removal of non-informative frames for wireless capsule endoscopy video segmentation," in *2012 IEEE International Conference on Automation and Logistics*, 2012, pp. 294–299.

[59] R. Sharma, R. Bhadu, S. K. Soni, and N. Varma, "Reduction of Redundant Frames in Active Wireless Capsule Endoscopy," 2019, pp. 1–7.

[60] Y. Fu, H. Liu, Y. Cheng, T. Yan, T. Li, and M. Q.-H. Meng, "Key-frame selection in WCE video based on shot detection," in *Proceedings of the 10th World Congress on Intelligent Control and Automation*, 2012, pp. 5030–5034.

[61] H.-G. Lee, M.-K. Choi, B.-S. Shin, and S.-C. Lee, "Reducing redundancy in wireless capsule endoscopy videos," *Comput. Biol. Med.*, vol. 43, no. 6, pp. 670–682, Jul. 2013.

[62] H. Liu, N. Pan, H. Lu, E. Song, Q. Wang, and C.-C. Hung, "Wireless Capsule Endoscopy Video Reduction Based on Camera Motion Estimation," *J. Digit. Imaging*, vol. 26, no. 2, pp. 287–301, Apr. 2013.

[63] Y. Chen, Y. Lan, and H. Ren, "Trimming the Wireless Capsule Endoscopic Video by Removing Redundant Frames," in *2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*, 2012, pp. 1–4.

[64] P. Vieira *et al.*, "Segmentation of small bowel tumor tissue in capsule endoscopy images by using the MAP algorithm.," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2012, pp. 4010–3, Jan. 2012.

[65] B. Li and M. Q.-H. Meng, "Wireless capsule endoscopy images enhancement via adaptive contrast diffusion," *J. Vis. Commun. Image Represent.*, vol. 23, no. 1, pp. 222–228, Jan. 2012.

[66] M. Ramaraj, S. Raghavan, and W. A. Khan, "Homomorphic filtering techniques for WCE image enhancement," in *2013 IEEE International Conference on Computational Intelligence and Computing Research*, 2013, pp. 1–5.

[67] H. Vu *et al.*, "Image-Enhanced Capsule Endoscopy Preserving the Original Color Tones," in *Abdominal Imaging. Computational and Clinical Applications*, H. Yoshida, G. Sakas, and M. G. Linguraru, Eds. Springer Berlin Heidelberg, 2012, pp. 35–43.

[68] H. Okuhata, H. Nakamura, S. Hara, H. Tsutsui, and T. Onoye, "Application of the real-time Retinex image enhancement for endoscopic images," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 3407–3410.

[69] M. S. Imtiaz and K. A. Wahid, "Image enhancement and space-variant color reproduction method for endoscopic images using adaptive sigmoid function," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 3905–3908.

[70] B. Jamlee Ludes and S. R. Norman, "Enhancement of Endoscopic Image Using TV-Image Decomposition," in *Proceedings of the International Conference on Soft Computing Systems*, L. P. Suresh and B. K. Panigrahi, Eds. Springer, 2016, pp. 67–75.

[71] M. Liedlgruber and A. Uhl, "Computer-Aided Decision Support Systems for Endoscopy in the Gastrointestinal Tract: A Review," *IEEE Rev. Biomed. Eng.*, vol. 4, pp. 73–88, 2011.

[72] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: challenges for essential progress," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 12, no. 3, pp. 172–186, Feb. 2015.

[73] S. Hwang, J. Oh, J. Cox, S. J. Tang, and H. F. Tibbals, "Blood Detection in Wireless Capsule Endoscopy using Expectation Maximization Clustering," in *Proc. SPIE 6144, Medical Imaging*, 2006, vol. 6144, no. 2006, p. 61441P.

[74] P. Y. Lau and P. L. Correia, "Detection of bleeding patterns in WCE video using multiple features," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 5601–5604.

[75] B. Giritharan, Xiaohui Yuan, Jianguo Liu, B. Buckles, JungHwan Oh, and Shou Jiang Tang, "Bleeding detection from capsule endoscopy videos," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 4780–4783.

[76] Y. S. Jung, Y. H. Kim, D. H. Lee, S. H. Lee, J. J. Song, and J. H. Kim, "Automatic patient-adaptive bleeding detection in a capsule endoscopy," 2009, p. 72603T.

[77] G. Pan, G. Yan, X. Qiu, and J. Cui, "Bleeding Detection in Wireless Capsule Endoscopy Based on Probabilistic Neural Network," *J. Med. Syst.*, vol. 35, no. 6, pp. 1477–1484, Dec. 2011.

[78] Guolan Lv, Guozheng Yan, and Zhiwu Wang, "Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 6643–6646.

[79] M. Hajabdollahi *et al.*, "Segmentation of bleeding regions in wireless capsule endoscopy for detection of informative frames," *Biomed. Signal Process. Control*, vol. 53, p. 101565, Aug. 2019.

[80] F. Deeba, S. K. Mohammed, F. M. Bui, and K. A. Wahid, "A Saliency-based Unsupervised Method for Angiectasia Detection in Endoscopic Video Frames," *J. Med. Biol. Eng.*, vol. 38, no. 2, pp. 325–335, Apr. 2018.

[81] F. Deeba, S. K. Mohammed, F. M. Bui, and K. A. Wahid, "A Saliency-Based Unsupervised Method for Angioectasia Detection in Capsule Endoscopic Images," in *The 39th Conference of The Canadian Medical and Biological Engineering*, 2016.

[82] F. Noya, M. A. Alvarez-Gonzalez, and R. Benitez, "Automated angiodysplasia detection from wireless capsule endoscopy," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3158–3161.

[83] K. Pogorelov *et al.*, "Deep learning and handcrafted feature based approaches for automatic detection of angiectasia," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2018, pp. 365–368.

[84] A. Shvets, V. Iglovikov, A. Rakhlin, and A. A. Kalinin, "Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks," *Cold Spring Harb. Lab.*, 2018.

[85] A. Tsuboi *et al.*, "Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images," *Dig. Endosc.*, p. den.13507, Oct. 2019.

[86] D. J. C. Barbosa, J. Ramos, and C. S. Lima, "Using Texture Analysis based on the Discrete Wavelet Transform," pp. 3012–3015, 2008.

[87] D. Barbosa, D. Roupar, J. Ramos, A. Tavares, and C. S. Lima, "Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images.," *Biomed. Eng. Online*, vol. 11, no. 1, p. 3, Jan. 2012.

[88] K. W. Abyoto, S. J. Wirdjosoedirdjo, and T. Watanabe, "Unsupervised Texture Segmentation Using Multiresolution Analysis for Feature Extraction," *J. Tokyo Univ. Inform. Sci.*, vol. 2, pp. 49–61, 1998.

[89] P. Szczypinski, R. Sriram, P. Sriram, and D. Reddy, "A model of deformable rings for interpretation of wireless capsule endoscopic videos," *Med. Image Anal.*, vol. 13, no. 2, pp. 312–324, 2009.

[90] C. S. Lima, D. Barbosa, J. Ramos, A. Tavares, L. Monteiro, and L. Carvalho, "Classification of endoscopic capsule images by using color wavelet features, higher order statistics and radial basis functions," *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 1242–1245, 2008.

[91] D. Barbosa, J. Ramos, and C. S. Lima, "Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2008, pp. 3012–5, Jan. 2008.

[92] D. Barbosa, J. Ramos, and C. S. Lima, "Wireless capsule endoscopic frame classification scheme based on higher order statistics of multi-scale texture descriptors," *4th Eur. Conf. Int. Fed. Med. Biol. Eng. - ECIFMBE 2008*, pp. 200–203, 2008.

[93] D. Barbosa, J. Ramos, A. Tavares, and C. S. Lima, "Detection of Small Bowel Tumors in Endoscopic Capsule Images by Modeling Non-Gaussianity of Texture Descriptors," *Int. J. Tomogr. Stat.*, vol. 14, no. S10, pp. 41–55, 2010.

[94] M. Alizadeh, O. H. Maghsoudi, K. Sharzehi, H. R. Hemati, A. K. Asl, and A. Talebpour, "Detection of small bowel tumor in wireless capsule endoscopyimages using an adaptive neuro-fuzzy inference system," *J. Biomed. Res.*, vol. 31, no. 5, pp. 419–427, Sep. 2017.

[95] P. Szczypiński, A. Klepaczko, M. Pazurek, and P. Daniel, "Texture and color based image segmentation and pathology detection in capsule endoscopy videos," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 396–411, Jan. 2014.

[96] G. Liu, G. Yan, S. Kuang, and Y. Wang, "Detection of small bowel tumor based on multi-scale curvelet analysis and fractal technology in capsule endoscopy," *Comput. Biol. Med.*, vol. 70, pp. 131–138, Mar. 2016.

[97] D. Y. Cheung, J. S. Kim, K.-N. Shim, and M.-G. Choi, "The Usefulness of Capsule Endoscopy for Small Bowel Tumors," *Clin. Endosc.*, vol. 49, no. 1, pp. 21–25, Jan. 2016.

[98] M. A. Ben Atitallah, R. Kachouri, M. Kammoun, and H. Mnif, "An efficient implementation of GLCM algorithm in FPGA," in *2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, 2018.

[99] A. Parvez and A. C. Phadke, "Efficient implementation of GLCM based texture feature computation using CUDA platform," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017.

[100] V. Prasath, "Polyp Detection and Segmentation from Video Capsule Endoscopy: A Review," *J. Imaging*, vol. 3, no. 1, p. 1, Dec. 2016.

[101] Baopu Li, M. Q.-H. Meng, and Lisheng Xu, "A comparative study of shape features for polyp detection in wireless capsule endoscopy images," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 3731–3734.

[102] B. Li, Y. Fan, M. Q.-H. Meng, and L. Qi, "Intestinal polyp recognition in capsule endoscopy images using color and shape features," in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2009, pp. 1490–1494.

[103] A. Karargyris and N. Bourbakis, "Identification of polyps in Wireless Capsule Endoscopy videos using Log Gabor filters," in *2009 IEEE/NIH Life Science Systems and Applications Workshop*, 2009, pp. 143–147.

[104] A. Karargyris and N. Bourbakis, "Detection of small bowel polyps and ulcers in wireless capsule

endoscopy videos.," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2777–86, Oct. 2011.

[105] Qian Zhao and M. Q.-H. Meng, "Polyp detection in wireless capsule endoscopy images using novel color texture features," in *2011 9th World Congress on Intelligent Control and Automation*, 2011, pp. 948–952.

[106] Q. Zhao, T. Dassopoulos, G. E. Mullin, M. Q.-H. Meng, and R. Kumar, "A Decision Fusion Strategy for Polyp Detection in Capsule Endoscopy," in *Studies in Health Technology and Informatics*, IOS Press, 2012, pp. 559–565.

[107] B. Li and M. Q.-H. Meng, "Automatic polyp detection for wireless capsule endoscopy images," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10952–10958, Sep. 2012.

[108] P. N. Figueiredo, I. N. Figueiredo, S. Prasath, and R. Tsai, "Automatic Polyp Detection in Pillcam Colon 2 Capsule Images and Videos: Preliminary Feasibility Report," *Diagn. Ther. Endosc.*, vol. 2011, pp. 1–7, May 2011.

[109] I. Figueiredo, S. Kumar, and P. Figueiredo, "An intelligent system for polyp detection in wireless capsule endoscopy images," in *Computational Vision and Medical Image Processing IV*, CRC Press, 2013, pp. 229–235.

[110] Y. Yuan and M. Q.-H. Meng, "Deep learning for polyp recognition in wireless capsule endoscopy images," *Med. Phys.*, vol. 44, no. 4, pp. 1379–1389, Apr. 2017.

[111] B. Li and M. Meng, "Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments," *Comput. Biol. Med.*, vol. 39, no. 2, pp. 141–147, 2009.

[112] B. Li and M. Q.-H. Meng, "Texture analysis for ulcer detection in capsule endoscopy images," *Image Vis. Comput.*, vol. 27, no. 9, pp. 1336–1342, Aug. 2009.

[113] L. Yu, P. C. Yuen, and J. Lai, "Ulcer Detection in Wireless Capsule Endoscopy Images," in *21st International Conference on Pattern Recognition (ICPR 2012)*, 2012.

[114] Y. Yuan, J. Wang, B. Li, and M. Q.-H. Meng, "Saliency Based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 2046–2057, Oct. 2015.

[115] S. Fan, L. Xu, Y. Fan, K. Wei, and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Phys. Med. Biol.*, vol. 63, no. 16, p. 165001, Aug. 2018.

[116] H. Alaskar, A. Hussain, N. Al-Aseem, P. Liatsis, and D. Al-Jumeily, "Application of Convolutional Neural Networks for Automated Ulcer Detection in Wireless Capsule Endoscopy Images," *Sensors*, vol. 19, no. 6, p. 1265, Mar. 2019.

[117] M. Hajabdollahi, R. Esfandiarpoor, E. Sabeti, N. Karimi, S. M. R. Soroushmehr, and S. Samavi, "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network," *Biomed. Signal Process. Control*, vol. 57, p. 101792, Mar. 2020.

[118] W. Yang, Y. Cao, Q. Zhao, Y. Ren, and Q. Liao, "Lesion Classification of Wireless Capsule Endoscopy Images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1238–1242.

[119] D. J. Bora, A. K. Gupta, and F. A. Khan, "Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 5, no. 2, pp. 192–203, 2015.

[120] D. Pascale, "A Review of RGB Color Spaces ... from xyY to R'G'B'." BabelColor, pp. 5–33, 2003.

[121] M. Horvath, "Isometric Land," 2008. [Online]. Available: http://isometricland.net/home.php. [Accessed: 20-Nov-2019].

[122] R. W. G. Hunt and M. R. Pointer, *Measuring Colour*. Chichester, UK: John Wiley & Sons, Ltd, 2011.

[123] A. Woodland, F. Labrosse, A. Woodland, and F. Labrosse, "On the separation of luminance from colour in images On the separation of luminance from colour in images," in *International Conference on Vision, Video and Graphics*, 2005, pp. 29–36.

[124] A. Ford and A. Roberts, "Colour Space Conversions," London, 1998.

[125] C. Poynton and B. Funt, "Perceptual uniformity in digital image representation and display," *Color Res. Appl.*, vol. 39, no. 1, pp. 6–15, Feb. 2014.

[126] P. J. Baldevbhai and R. S. Anand, "Color Image Segmentation for Medical Images using L * a * b * Color Space," *J. Electron. Commun. Eng.*, vol. 1, no. 2, pp. 24–45, 2012.

[127] S. Bansal and D. Aggarwal, "Color Image Segmentation using CIELab Color Space using Ant Colony Optimization," *Int. J. Comput. Appl.*, vol. 29, no. 9, pp. 28–34, 2011.
[128] P. M. Vieira, "Segmentação de imagens da Cápsula Endoscópica com aplicações na deteção de tumores e na reconstrução 3-D da mucosa intestinal," Universidade do Minho, 2013.
[129] T. Mouw, "Tolerancing Part 3: Color Space vs. Color Tolerance," *x-Rite*, 2018. [Online]. Available: https://www.xrite.com/blog/tolerancing-part-3. [Accessed: 20-Nov-2019].
[130] D. Srivastava, R. Wadhvani, and M. Gyanchandani, "A Review : Color Feature Extraction Methods for Content Based Image Retrieval," *IJCEM Int. J. Comput. Eng. Manag.*, vol. 18, no. 3, pp. 9–13, 2015.
[131] G. N. Srinivasan and G. Shobha, "Statistical Texture Analysis," *Proc. World Acad. Sci. Eng. Technol.*, vol. 36, pp. 1264–1269, 2008.
[132] M. S. Banu and K. Nallaperumal, "Analysis of Color Feature Extraction Techniques for Pathology Image Retrieval System," in *Computational Intelligence and Computing Research (ICCIC)*, 2010.
[133] J.-M. Marin, K. Mengersen, and C. P. Robert, "Bayesian Modelling and Inference on Mixtures of Distributions," in *Handbook of Statistics*, D. K. Dey and C. R. Rao, Eds. New York, USA: Springer-Verlag, 2005, pp. 459–507.
[134] D. J. C. Mackay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, Massachusetts: Cambridge University Press, 2003.
[135] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," *Int. J. Comput. Sci. Inst.*, vol. 1198, no. 510, 1998.
[136] S. Borman, "The Expectation Maximization Algorithm: A short tutorial," 2004.
[137] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[138] H. F. Walker and P. Ni, "Anderson Acceleration for Fixed-Point Iterations," *SIAM J. Numer. Anal.*, vol. 49, no. 4, pp. 1715–1735, Jan. 2011.
[139] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain.," *IEEE Trans. Med. Imaging*, vol. 18, pp. 897–908, 1999.
[140] S. Marsland, *Machine Learning An algorithmic Perspective*, 2nd ed. CRC Press, 2015.
[141] S. G. Soares, "Ensemble Learning Methodologies for Soft Sensor Development in Industrial Processes," University of Coimbra, 2015.
[142] B. Li and M. Q. H. Meng, "Computer-aided detection of bleeding regions for capsule endoscopy images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1032–1039, 2009.
[143] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. 2014.
[144] Y. Lan, Y. C. Soh, and G.-B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing*, vol. 72, no. 13–15, pp. 3391–3395, Aug. 2009.
[145] Y. Lv, J. Liu, T. Yang, and D. Zeng, "A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler," *Energy*, vol. 55, pp. 319–329, Jun. 2013.
[146] M. I. Jordan and L. Xu, "Convergence results for the EM approach to mixtures of experts architectures," *Neural Networks*, vol. 8, no. 9, pp. 1409–1431, Jan. 1995.
[147] L. Deng, "Deep Learning: Methods and Applications," *Found. Trends® Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.
[148] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[149] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, May 2016.
[150] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *The 2011 International Joint Conference on Neural Networks*, 2011, pp. 1918–1921.
[151] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, Big, Simple Neural Nets for Handwritten Digit Recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, Dec. 2010.
[152] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*,

vol. 115, no. 3, pp. 211–252, Dec. 2015.

[153] A. Canziani, A. Paszke, and E. Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications," May 2016.

[154] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[155] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.

[156] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[157] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[158] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, "Simple Does It: Weakly Supervised Instance and Semantic Segmentation," Mar. 2016.

[159] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[160] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask Scoring R-CNN," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6402–6411.

[161] T. Cheng, X. Wang, L. Huang, and W. Liu, "Boundary-preserving Mask R-CNN," in *ECCV 2020*, 2020.

[162] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8759–8768.

[163] N. R. Freitas, P.M.Vieira, A. I. Vaz, and C. S. Lima, *Stochastic Optimization by using Higher-Order Moments*. Manuscript Submitted for publication, 2020.

[164] Y. Nesterov, "A Method for Solving a Convex Programming Problem with Convergence Rate O(1/K^2)," *Sov. Math. Dokl.*, vol. 27, no. 2, pp. 372–376, 1983.

[165] D. K. Iakovidis and A. Koulaouzidis, "Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software," *Gastrointest. Endosc.*, vol. 80, no. 5, pp. 877–883, Nov. 2014.

[166] D. K. Iakovidis and A. Koulaouzidis, "Automatic lesion detection in wireless capsule endoscopy - A simple solution for a complex problem," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2236–2240.

[167] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: challenges for essential progress," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 12, no. 3, pp. 172–186, Mar. 2015.

[168] A. Koulaouzidis and D. K. Iakovidis, "KID: Koulaouzidis-Iakovidis Database for Capsule Endoscopy." [Online]. Available: http://is-innovation.eu/kid.

[169] P. M. Vieira, J. Ramos, and C. S. Lima, "Automatic detection of small bowel tumors in endoscopic capsule images by ROI selection based on discarded lightness information," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 3025–3028.

[170] P. Harrington, *Machine Learning in Action*. Greenwich, CT, USA: Manning Publications, 2012.

[171] D. J. C. Barbosa, J. Ramos, J. H. Correia, and C. S. Lima, "Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors.," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2009, vol. 2009, pp. 6683–6.

[172] A. Koulaouzidis *et al.*, "KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes," *Endosc. Int. Open*, vol. 05, no. 06, pp. E477–E483, Jun. 2017.

[173] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, and V. P. Plagianakos, "Detecting and Locating Gastrointestinal Anomalies Using Deep Learning and Iterative Cluster Unification," *IEEE Trans. Med. Imaging*, vol. 37, no. 10, pp. 2196–2210, Oct. 2018.

[174] K. McLaren, "XIII-The Development of the CIE 1976 (L* a* b*) Uniform Colour Space and Colour-difference Formula," *J. Soc. Dye. Colour.*, vol. 92, no. 9, pp. 338–341, Oct. 2008.

[175] J. Cotter *et al.*, "Virtual chromoendoscopy in small bowel capsule endoscopy: New light or a cast of shadow?," *World J. Gastrointest. Endosc.*, vol. 6, no. 8, p. 359, 2014.

[176] P.-N. D'Halluin *et al.*, "Does the 'Suspected Blood Indicator' improve the detection of bleeding lesions by capsule endoscopy?," *Gastrointest. Endosc.*, vol. 61, no. 2, pp. 243–9, Mar. 2005.

[177] J. M. Buscaglia *et al.*, "Performance Characteristics of the Suspected Blood Indicator Feature in Capsule Endoscopy According to Indication for Study," *Clin. Gastroenterol. Hepatol.*, vol. 6, no. 3, pp. 298–301, 2008.

[178] H. Imagawa *et al.*, "Improved visibility of lesions of the small intestine via capsule endoscopy with computed virtual chromoendoscopy," *Gastrointest. Endosc.*, vol. 73, no. 2, pp. 299–306, Feb. 2011.

[179] C. Arieira *et al.*, "Su1259 IS THE SOFTWARE TOP 100 A RELIABLE TOOL IN CAPSULE ENDOSCOPY?," *Gastrointest. Endosc.*, vol. 87, no. 6, p. AB301, Jun. 2018.

[180] P. Boal Carvalho *et al.*, "Virtual chromoendoscopy improves the diagnostic yield of small bowel capsule endoscopy in obscure gastrointestinal bleeding," *Dig. Liver Dis.*, vol. 48, no. 2, pp. 172–175, Feb. 2016.

[181] A. J. Postgate, A. Fitzpatrick, P. Bassett, and C. H. Fraser, "Polyp Detection and Size Estimation At Capsule Endoscopy-Does Experience Improve Accuracy? A Prospective Animal-Model Study," *Gastrointest. Endosc.*, vol. 65, no. 5, p. AB316, Apr. 2007.

[182] GlobalData, "Capsule Endoscopes – Medical Devices Pipeline Assessment," London, 2020.

[183] R. Leenhardt *et al.*, "CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy," *Endosc. Int. Open*, vol. 08, no. 03, pp. E415–E420, Mar. 2020.