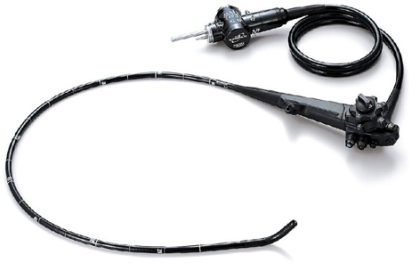
Computer-aided Classification of Endoscopic Images from the Gastrointestinal Tract

by
Michael Liedlgruber

Cumulative thesis submitted to the
Faculty of Natural Sciences, University of Salzburg
in partial fulfillment of the requirements
for the Doctoral Degree.

Submitted May 2011



Thesis Supervisor
Ao. Prof. Dr. Mag. rer. nat. Andreas Uhl
Department of Computer Sciences
University of Salzburg, Salzburg, Austria

Abstract

In modern medicine endoscopy plays a very important role as it allows physicians to detect severe diseases in early development stages already. Especially the gastrointestinal tract is examined routinely in order to detect pre-malignant and possibly malignant diseases. As a consequence, the mortality rate for many diseases, especially different types of cancers, has been lowered drastically throughout the last years.

In this cumulative thesis we specifically focus on computer-aided decision support systems targeting the gastrointestinal tract as those may potentially help to save time, lower the cost of endoscopic procedures, and lower the risk of such procedures for the patients. To underpin the importance of such systems this thesis includes a detailed overview and discussion of methods published within the two past decades, targeting this field of research.

Furthermore, we evaluate different types of features and classifiers throughout this work, in order to obtain a robust system for the classification of endoscopic images. The features used throughout this work either investigate local texture properties, are based on statistical texture features, or aim at a shape-based classification.

We also investigate the use of an ensemble classifier, which combines the outcome of different weak methods in order to stabilize the results of our methods across the different images classes.

To show the feasibility of the methods proposed we carry out numerous experiments on two distinct image databases. One database contains endoscopic images showing polyps from the colon, which we aim to classify correctly. The second database contains endoscopic imagery from celiac disease patients. In the case of the latter database we aim to detect the presence of celiac disease.

In order to assess real-time capabilities of our methods we also carry out performance tests and discuss the respective results.

In der modernen Medizin spielt die Endoskopie eine wichtige Rolle, da diese Untersuchungsmethode einem Mediziner die Möglichkeit gibt, schwerwiegende Erkrankungen frühzeitig zu erkennen. Vor allem der Gastrointestinaltrakt wird routinemäßig untersucht, um eventuell bösartige Veränderungen im Darm oder Vorstufen davon rechtzeitig erkennen zu können. Im Laufe der letzten Jahre ermöglichte dies eine drastische Senkung der Sterblichkeitsrate bei vielen verschiedenen Krankheitsbildern – vor allem bei Darmkrebs.

In dieser kumulativen Dissertation widmen wir uns im Speziellen Computer-gestützten Entscheidungssystemen, welche auf die Untersuchung des Gastrointestinaltrakts abzielen. Derartige Systeme können in Zukunft dabei helfen, den Zeitaufwand für Endoskopieuntersuchungen und die damit verbundenen Risiken für Patienten und Untersuchungskosten zu minimieren. Um die Wichtigkeit derartiger Systemen zu belegen, beinhaltet diese Dissertation eine ausführliche Zusammenfassung und Diskussion von Systemen, welche in den letzten zwei Jahrzehnten entwickelt wurden.

Des Weiteren evaluieren wir in dieser Arbeit verschiedene Typen von Merkmalen und Klassifikationsalgorithmen, mit dem Ziel eine robuste Klassifizierung von Endoskopieaufnahmen zu erreichen. Die dabei verwendeten Merkmale beschreiben Bilder auf verschiedene Art und Weise. Während manche Merkmale auf lokalen Bildeigenschaften oder statistischen Textureigenschaften basieren, zielen andere auf die Beschreibung vom markanten Strukturen in Endoskopieaufnahmen ab.

Um die Ergebnisse zu verbessern und über mehrere Bildklassen hinweg zu stabilisieren, evaluieren wir zudem einen Ensemble-Klassifikationsalgorithmus. Dieser kombiniert die Ausgaben von verschiedenen Methoden, um bei der Bildklassifikation noch genauere und stabilere Ergebnisse zu erreichen.

Um zu prüfen, ob die vorgestellten Methoden im Rahmen einer Klassifikation von Endoskopieaufnahmen erfolgreich sind, führen wir zahlreiche Experimente auf Basis von zwei verschiedenen Bilddatenbanken durch. Die erste Datenbank besteht aus Bildern von Polypen im Dickdarm, welche von den vorgestellten Methoden unterschieden werden sollen. Im Fall der zweiten Datenbank, welche aus Endoskopieaufnahmen von Zöliakie-Patienten besteht, ist das Ziel die Erkennung von Zöliakie.

Um auch Aussagen über die Echtzeitfähigkeit unserer Methoden treffen zu können, führen wir zudem Geschwindigkeitsmessungen bei den vorgestellten Methoden durch und diskutieren die entsprechenden Ergebnisse.

Acknowledgments

> It is good to have an end to journey toward, but it is the journey that matters in the end.
>
> *- Ursula K. LeGuin*

Writing this thesis sometimes was a very challenging task. But after all it was a very interesting experience which allowed me to delve more into the realm of pattern recognition. Nevertheless, this thesis would not have been possible without the support and patience of certain people.

Hence, special thanks goes to my advisor Andreas Uhl for his great support and keeping me on track throughout the past years.

Furthermore I want to thank my family which always believed in me and supported me during my study. I also want to thank my girlfriend for her patience when I again and again couldn't resist talking about my thesis.

I also want to thank the medical experts who contributed to this thesis by providing image material and histopathological results, in particular Michael Häfner and Andreas Vécsei.

I also like to thank my colleagues from the WaveLab working group at our institute for their support and many inspiring discussions and Peter Meerwald for providing me with his PhD thesis LaTeX template, which indeed saved me a lot of headache.

Last but not least I'd like to thank the Austrian Science Fund (FWF) and the Austrian National Bank "Jubiläumsfonds" for funding this thesis.

Michael Liedlgruber
Salzburg, May 2011

Contents

1. Introduction

1.1. Medical and Technical Background

Since medical endoscopy is a minimally invasive and relatively painless procedure, allowing to inspect the inner cavities of the human body, endoscopes play an important role in modern medicine. In clinical practice different organs such as the respiratory tract, the urinary tract, and the female reproductive system are regularly inspected by using an endoscope. Another important field in medical endoscopy, which this work focuses at, is the inspection of the gastrointestinal tract (GI tract). The parts of the human GI tract, which are most commonly inspected with an endoscope, are shown in Figure 1.1.

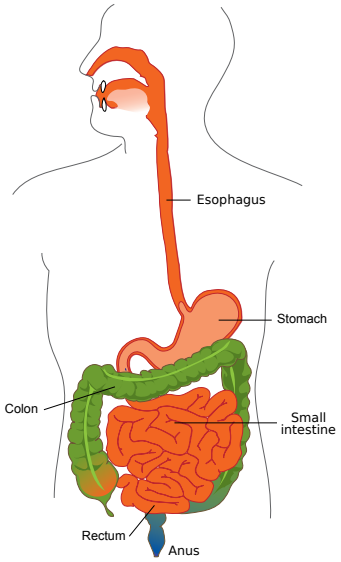


Figure 1.1.: A schematic illustration of the human GI tract.

Based on endoscopy of the GI tract, physicians are able to detect severe diseases already in early development stages and therefore the mortality rate for many diseases, especially different types of cancers, has been lowered drastically throughout the last years. Some examples of conditions which are known to be premalignant or to increase the risk of cancer in the GI tract are adenomas, Barrett's esophagus, Crohn's disease, celiac disease, GI bleeding, and a Helicobacter pylori infection.

In general the area of applications for endoscopes is wide. Besides medical procedures, endoscopes are also used to inspect airplane turbines, pipes in buildings or industrial machinery, car engines, tanks in ships, and for veterinary endoscopy. However, throughout this work the terms "endoscope" and "endoscopy" always denote the medical device and procedure, respectively.

Endoscopy, as we know it today, is performed using a flexible endoscope, sometimes also referred to as videoscope. This type of endoscope has been introduced in the mid 1960s. While the first endoscopes used fiber optics and an eyepiece lens to visualize the inner cavities of the human body, modern endoscopes are very compact devices, including a light source, and a CCD or CMOS chip for taking pictures. But the basic concept did not change very much since those days. In addition to the digital imaging chip, modern endoscopes contain a light source at the distal tip and are equipped with an accessory channel, which allows the entry of medical instruments for example to take tissue samples, perform cleansing of poorly prepared areas, perform polypectomies, and perform endoscopic resections without any invasive surgery involved.

Due to the digital imaging chips used in modern endoscopes such devices are also regularly used to take digital pictures and record video sequences. These abilities created the whole new field of computer-aided decision support systems (CADSSs) in medical endoscopy. The aim of such systems is to predict pathologies and thus to assist a medical expert in improving the

accuracy of medical diagnosis [18]. Hence, the development of such systems is motivated by the following key aspects:

- **Saving time and reducing cost**
  CADSSs usually help to identify regions which may be of particular interest for a medical expert. Since, in general, such systems are designed to detect abnormal changes within the GI tract (e.g. neoplastic or metaplastic changes) they are also able to help avoiding possibly unnecessary biopsies, allowing to perform more targeted biopsies. In the course of such targeted biopsies the time needed for an endoscopy procedure may be lowered drastically. As a consequence the procedure gets more comfortable for the patient. Furthermore, since the number of necessary biopsies to be taken can be reduced, the time needed for the subsequent histopathologic examination can be reduced too.

  Reducing the time needed for an endoscopy and the subsequent histopathologic examination also results in a reduction of costs associated with such procedures.

  Considering the fact that re-investigating a video recorded during an endoscopy session may consume as much time as the endoscopic procedure itself, there is a potential to save time and costs if CADSSs are able to identify parts of such videos which might be of interest for a medical expert.

- **Enhancing accuracy of diagnosis**
  Endoscopy is a tedious procedure, demanding a constantly high level of concentration by an endoscopist. This is mainly due to the fact that missing an abnormality during endoscopy may be hazardous. Especially lesions which are rather small or only noticeable for a short fraction of time may be missed easily. This particularly applies to wireless capsule endoscopy (WCE) [17], where a lesion may show up in a single frame only (out of more than 50 000 frames in total!). This is due to the rather low frame rate of currently available capsules which usually is about two frames per second.

  The advantage of CADSSs is that computers always exhibit the same level of "concentration" – no matter how long an endoscopic procedure lasts or how many of such procedures a computer has to analyze in series. Also situations with bad light conditions, poor image quality, or poor contrast usually pose a problem for a medical expert, increasing the risk of missed lesions. If designed properly, a computer-based systems may be able to cope well with such circumstances. The constant level of "concentration" and the resistance against poor image conditions may help to avoid missing lesions, leading to an enhanced diagnostic accuracy of an endoscopic procedure.

  But, as indicated in [14], sometimes there is a more simple reason for missed lesions: an abnormality may be simply get missed since it is misinterpreted and therefore not recognized as being abnormal. The likelihood of missing a lesion for this reason heavily depends on the expertise of the medical expert who performs the endoscopy, leading to a lowered inter-observer agreement level. CADSSs, on the other hand, do not suffer from this problem. If the detection of abnormalities is designed in a robust way and there is sufficient training data available, a CADSS is expected to always deliver roughly the same level of accuracy.

- **Training of experts to new endoscopic imaging modalities**
  Throughout the past few years there have been many major advances in endoscopy. While the new imaging modalities have the potential to greatly increase the efficiency of endoscopy, medical experts need to get trained on these new techniques. In order to assess

the skills of a medical expert on new techniques, CADSSs can be used as an expert training tool to predict pathology, verify the detection or prediction performance of a medical expert, and serve as an educational resource.

As a consequence, a rising interest in the field of CADSSs targeted at endoscopy in the GI tract can be observed within approximately the past two decades [71].

1.2. Different Types of CADSSs

Basically CADSSs can be categorized into medical image classification (MIC) systems and systems which are based on content-based image retrieval (CBIR) strategies. While these two concepts share some basic building blocks, there are also fundamental differences [70]. Classifier-based approaches provide a second opinion to a medical expert by predicting the pathology in a fully autonomous manner for an unknown image with unknown pathology. CBIR-based systems, on the other hand, can be regarded as the digital counterpart of current clinical practice, where medical experts compare cases with unknown pathology to cases with an already known and verified diagnosis. Based on a suitable metric the system returns a list of images which are most similar to the image with unknown pathology. Usually the outcome of such systems can then be refined by a medical expert by issuing a new image query based on one or more images which seem to be relevant to the medical expert (commonly termed "relevance feedback").

One of the main differences between the different system types is therefore the fact that, while MIC-systems perform the prediction in a fully autonomous fashion, CBIR-based systems allow a medical expert to refine the prediction process. In case of CBIR, it is up to the medical expert to choose the final match (along with the respective known pathology) and therefore make a judgment on the pathology of the unknown image.

While most of the work found in literature is of MIC-type there also exist a few CBIR-based approaches (e.g. [6, 79]). The dominance of MIC systems may be attributed to the fact that such systems are potentially suited for an online diagnosis (performed in real-time during the medical examination already). CBIR systems, on the other hand, are restricted to offline scenarios mainly due to the interactive nature of such systems.

In addition, CADSSs targeting at endoscopy in the GI tract can also be divided into systems which aim at the detection of abnormalities (e.g. [46, 25]) and systems which go one steps further and provide a prediction about the underlying pathology (e.g. [77, 95, 15]). In case of the latter type abnormalities of interest are either detected implicitly or abnormality candidates are found using a different method and used as input to the classification system.

1.3. GI Tract Parts and Pathologies Under Investigation

Despite the fact that there exist different types of systems throughout literature, these approaches also differ in terms of the part of the GI tract they are targeted at. As a consequence there exists a variety of pathologies which are targeted by different CADSSs.

The vast majority of approaches is targeted at the detection of classification of colonic polyps. This can be explained by the high prevalence of such polyps and the fact that these may develop into colon cancer [71]. Systems aiming at the detection or classification of polyps in the colon have been proposed for example in [77, 46, 95, 39, 54, 48, 29, 25].

Another rather high share of CADSSs-related research aims at distinguishing between normal and abnormal regions without being specific about the underlying pathology. Besides the detection of such regions in the colon (e.g. [96, 75]), also the small bowel is targeted by some

work [10, 56]. In addition there exist quite a few approaches which do not focus on a particular GI tract part but examine all parts of the GI tract using WCE (e.g. [100, 67, 59]).

Another prominent field of research is the detection of gastrointestinal bleeding since this pathology may be an indication for many diseases such as, for example, colon cancer, Crohn's disease, esophageal cancer, small intestine cancer, or the typhoid fever. Since GI bleeding may occur in many different parts of the GI tract it is no surprise that related approaches are usually based on WCE (e.g. [64, 23, 49, 51, 65, 84, 83, 1]).

The remaining work found in literature targets at the detection or classification of other pathologies such as ulcers (e.g. [66, 92]), celiac disease (e.g. [99, 98, 97]), tumors (e.g. [8]), Crohn's disease (e.g. [9, 22]), cancer (e.g. [90]), intestinal dysfunctions (e.g. [87]), Barrett's esophagus (e.g. [79]), or Helicobacter pylori (e.g. [45]).

1.4. Different Types of Features

In order to detect or classify different pathologies within the GI tract various different types of features are commonly used throughout the respective literature. These features can be roughly grouped into low-level features and high-level features.

1.4.1. Low-level features

Despite the fact that the low-level features presented below all aim at describing textural content within images, they can be further subdivided into spatial domain features and frequency domain features.

Spatial domain features

Features in this category are extracted directly in the spatial domain and usually based on the color information stored in color channels.

The simplest approaches in this category use information such as the pixel color [83] or a combination of pixel colors with the respective pixel positions [2] with a suitable classifier in order to perform a classification. Other rather simple approaches just count pixels of certain colors or color tones and decide upon eventual pathologies by thresholding the pixel counter [1, 64].

Statistical measures such as for example the mean value or the standard deviation computed from pixel colors or grayscale images are also frequently used features (e.g. [15, 45, 59, 87]).

Histograms of color channels or images are also features which are found very often throughout literature. Usually either histograms are computed and used directly for a similarity-based classification (e.g. [38, 37, 67, 23]) or features like Haralick features [43] or other statistical features are extracted from histograms and used for a subsequent classification (e.g. [8, 75, 22, 57, 99, 100]). Sometimes the histogram features are complemented by other features to increase the classification performance (e.g. [23]).

Other features, which can be found in literature dealing with CADSSs targeting at medical endoscopy of the GI tract, are based on operators aiming at capturing local texture properties. Such operators are the Local Binary Patterns (LBP) operator [82] and the Texture spectrum (TS) operator, originally proposed in [101]. After one of these operators has been applied to an image, usually histograms (e.g. [3, 47]) or statistical features based on the transformed images or histograms (e.g. [58, 96, 100, 47]) are extracted and used for classification. In order to capture local texture features across an image sometimes the Scale Invariant Feature Transform (SIFT) [74] is used to extract scale invariant features from endoscopic images (e.g. [6, 4, 5, 93]).

Frequency domain features

Features falling into this category are extracted from an image or color channel after applying some sort of transformation of the data into the frequency domain.

Throughout literature dealing with CADSSs targeted at endoscopy of the GI tract the Discrete Wavelet Transform (DWT) [76] has been established and is used quite frequently to obtain image features. Variants of the DWT, such as the Dual-tree Complex Wavelet Transform (DT-CWT) [88], the Curvelet transform [12], or the Stationary Wavelet Transform (SWT) [85], have also been used to obtain approximate shift-invariant and rotation-invariant features. While some approaches construct co-occurrence matrices from wavelet subbands and extract Haralick features from them (e.g. [53, 48, 73, 77, 56, 54]), other approaches extract statistical features from wavelet subbands (e.g. [45, 40, 99, 63]), approximate coefficient distributions and use the estimated distribution parameters for the classification (e.g. [62, 28]), or apply another transformation on the wavelet subbands in order to obtain features (e.g. [66]).

Despite the undisputed dominance of wavelet features, there exists work which uses other frequency transforms such as the Fast Fourier Transform (FFT) or the Discrete Cosine Transform (DCT). These approaches usually extract statistical features from the transformation coefficients in order to obtain features for the classification (e.g. [30, 31, 29, 30]).

1.4.2. High-level features

Features in this category usually describe geometrical properties of shapes extracted from images. Thus, instead of describing textural properties within an image high-level features describe an image in a more abstract way. Hence, usually some sort of edge detection algorithm like the Canny edge detector [13] or the SUSAN edge detector [89] is used to obtain the edge information for an image. Based on such an edge image features different features describing shapes may be used for the classification of endoscopic images (e.g. [25, 26, 95, 50, 52, 91, 60, 79]).

2. Contribution

Our work published throughout the past years can be divided into the following three categories: overview articles, methods for the classification of colonic polyps, and methods aiming at the detection of celiac disease.

2.1. Overview Articles

Literature with respect to computer-aided decision support systems targeted at endoscopy of the GI tract is highly dispersed. Hence, in the past we published an introductory paper, a technical report, and a book chapter, which summarize and discuss different aspects of this research topic.

While in [69] we just provide a brief summary of CADSSs targeted at endoscopic images from the GI tract, the technical report [71] gives a comprehensive overview of research related to the this topic. Besides presenting facts and figures concerning related work, we also critically discuss different issues which are inherent to this topic of research (e.g. lack of common image databases, different evaluation protocols used throughout literature, performance issues, obtaining ground truth information, and comparison of system accuracies among different publications).

The chapter published in [70] more specifically explains the different steps which are common to pathology prediction systems (e.g., obtaining ground truth information, assembling training image databases, the advantages and disadvantages of different validation protocols usually used throughout related literature). In addition, this work elaborates on the differences between CBIR and MIC systems and the respective similarities and differences with respect to the different processing steps in such prediction systems. This is accompanied by a comparison of example approaches from literature.

Publications (sorted chronologically)

[69] M. Liedlgruber and A. Uhl. Endoscopic image processing - an overview. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 707–712, Salzburg, Austria, Sept. 2009

[70] M. Liedlgruber and A. Uhl. Predicting pathology in medical decision support systems in endoscopy of the gastrointestinal tract. In C. Jao, editor, *Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain*, pages 195–214. InTech, Rijeka, Croatia, 2011

[71] M. Liedlgruber and A. Uhl. A summary of research targeted at computer-aided decision support in endoscopy of the gastrointestinal tract. Technical Report 2011-01, Department of Computer Sciences, University of Salzburg, Austria, http://www.cosy.sbg.ac.at/research/tr.html, 2011

2.2. Polyp Classification

2.2.1. Medical background

Due to the fact that colonic polyps have a rather high prevalence and are known to either develop into cancer or to be precursors of colon cancer, an early detection of such pathologies can lower the mortality rate drastically. Hence, throughout the past years we developed various methods aiming at an assessment of the malignant potential of colonic polyps with the aim of avoiding random and, probably, unnecessary biopsies. As a consequence such systems could potentially help to save time, lower the cost for colonoscopy procedures, and reduce the risk of complications during such procedures.

In order to be able to distinguish between the different types of polyps we based our work on the pit pattern classification scheme, originally reported by Kudo et al. [61]. Based on the visual pattern of the mucosal surface this system allows to differentiate between normal mucosa, hyperplastic lesions (non-neoplastic), adenomas (a pre-malignant condition), and malignant cancer. Thus this classification it is also a convenient choice for an automated image classification. Since the pit pattern classification distinguishes between six different pit pattern types, which can be grouped into non-neoplastic and neoplastic lesions, a 2-class classification is possible as well as a 6-class classification.

The images used throughout our publications have been obtained using a zoom-colonoscope with a magnification factor of 150. In addition, topical staining has been applied in order to visually enhance mucosal crypt patterns or vascular features.

2.2.2. Our contribution

As already mentioned in Section 1.4, there exists a variety of different features which might potentially be used for the classification of endoscopic images. Hence, in the past we evaluated different types of features in terms of their discriminative power with respect to polyp classification.

In [42] we compared the Local Discriminant Bases (LDB) algorithm [86], which has been developed with discrimination between different image classes in mind, against the pyramidal DWT in terms of the classification rates, using statistical features based on wavelet coefficients. We showed, that the LDB algorithm is able to deliver higher classification accuracies as compared to the DWT.

We then presented a larger study in [68] by comparing the classification accuracies of different types of wavelet-based operators. The main focus of this work, however, was a comparison of statistical wavelet features and structural features based on the decomposition structures obtained by the Best-Basis algorithm [16]. The outcome of this work was that in most cases the statistical features deliver superior classification rates as compared to the structural features.

The impact of combining features from different colors channels on classification accuracies has been investigated in more detail in [34]. For this purpose we compared the classification rates of previously developed wavelet-based features, histogram-based features [38], and Fourier-based features [31] when extracted from grayscale images or multiple color channels. We were able to show that extracting features from all color channels available constantly improved the classification rates of the methods compared. This observation lead to the decision that we exploited color information where possible in all subsequent publications.

In addition, we introduced an ensemble classifier in [34] which combines the prediction outcomes of different methods in order to improve and stabilize the overall classification rate and the classification rates across the different image classes. We showed, that, at least in case of

the more challenging 6-classes case, the ensemble classifier is able to considerably improve the classification rates.

In [32] we combined the concept of Gaussian Markov Random Fields (GMRF) with the wavelet transform since wavelet-based features already delivered very promising classification results in previous work. For this purpose we estimated Markov parameters from subbands resulting from a DWT. In addition we introduced directional GMRF neighborhoods specifically tailored to the wavelet domain, exploiting the different directions of the different detail subbands. We were able to show that moving the Markov parameter estimation to the wavelet domain considerably improves the classification rates. The introduction of the directional neighborhoods improved the classification rates even more.

In a more recent work we also evaluated a noise-robust version of the LBP operator for a classification of polyps [33]. We showed that, compared to the original LBP operator, our extension is able to deliver higher classification accuracies, especially in the 6-classes case. Since we combined information from different color channels into 2D-histograms this method suffered in terms of the time needed for image classification. In order to overcome this limitation, we proposed a novel texture operator based on LBP in a more recent work [41]. With the goal of developing a compact and fast operator we were able to show that, while delivering roughly the same classification accuracies, our new operator is up to 7.5 times faster compared to the method proposed in [33].

Due to the visual nature of the pit pattern classification we also investigated shape-based features. In [35] we classified the pit pattern images based on the density of pit candidates found within the images. While the results achieved were very promising, the density of pit candidates is only suitable for a 2-class classification. To allow a 6-class classification too we investigated the combination of shape-based and texture-based features in a follow-up work in [36]. This combination yielded highly competitive results as compared to previously developed approaches.

Publications (sorted chronologically)

[42] M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using wavelet texture features. In W. Sandham, D. Hamilton, and C. James, editors, *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06)*, pages 1–4, Glasgow, Scotland, UK, July 2006

[68] M. Liedlgruber and A. Uhl. Statistical and structural wavelet packet features for Pit pattern classification in zoom-endoscopic colon images. In P. Dondon, V. Mladenov, S. Impedovo, and S. Cepisca, editors, *Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07)*, pages 147–152, Arcachon, France, Oct. 2007

[34] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using multichannel features and multiclassification. In D. F. T.P. Exarchos, A. Papadopoulos, editor, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, pages 335–350. IGI Global, Hershey, PA, USA, 2009

[32] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Combining Gaussian Markov random fields with the discrete wavelet transform for endoscopic image classification. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'09)*, pages 177–182, Santorini, Greece, July 2009

[33] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using extended local binary patterns. In *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, pages 1–4, Larnaca, Cyprus, Nov. 2009

[35] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Classification of endoscopic images using Delaunay triangulation-based edge features. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, volume 6112 of *Springer LNCS*, pages 131–140, Povoa de Varzim, Portugal, June 2010

[36] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Endoscopic image classification using edge-based features. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, pages 2724–2727, Istanbul, Turkey, Aug. 2010

[41] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Medical Image Analysis*, 16(1):75–86, 2012

2.3. Celiac Disease Detection

2.3.1. Medical background

Celiac disease, commonly known as gluten intolerance, is a complex autoimmune disorder that affects the small bowel in genetically predisposed individuals of all age groups after introduction of food containing gluten. Characteristic for the disease is an inflammatory reaction in the mucosa of the small intestine. During the course of the disease the mucosa looses its absorptive villi and hyperplasia of the enteric crypts occurs, leading to a diminished ability to absorb nutrients.

Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. During standard upper endoscopy at least four duodenal biopsies are taken. Microscopic changes within these specimen are then classified in a histological analysis according to the Marsh classification proposed in 1992 [78]. Subsequently, Oberhuber et al. proposed the modified Marsh classification [80] which distinguishes between classes Marsh-0 to Marsh-3, with subclasses Marsh-3a, Marsh-3b, and Marsh-3c, resulting in a total number of six classes.

According to the modified Marsh classification Marsh-0 denotes a healthy mucosa (without visible changes of the villous structure) and Marsh-3c designates a complete absence of villi (villous atrophy). Since visible changes in the villi structure can be observed only between classes Marsh-0 and classes Marsh-3a to Marsh-3c, our work either carries out a 2-class classification or a 4-class classification.

2.3.2. Our contribution

In [98] we evaluated the Fourier features from [31] and the ensemble classifier proposed in [34] on celiac disease images. We showed that the results achieved were very promising. In addition our ensemble classifier slightly improved the classification accuracies as compared to the single methods.

In [44] we aimed at assessing the impact of different endoscopic image capturing techniques (i.e. classical method and modified immersion technique) on an automated classification of celiac disease imagery. For this evaluation we compared the results achieved by previously

developed methods on different image databases (e.g. [98, 32, 42]). We were able to show that the capturing technique used has indeed an impact on the classification accuracies achieved.

An image degradation commonly seen in endoscopic images are barrel-type distortions. Since these distortions are claimed to affect diagnosis [11], we also investigated the impact of such distortions on the accuracy of an automated diagnosis. While in [27] we investigated the impact of distortion correction only, we examined the effect of distortions as well as distortion correction in a more recent work [72]. In addition we provide a more detailed statistical analysis of an eventual effect of distortions and distortion correction in [72]. From the results obtained in [27] we were able to show that distortion correction does not automatically lead to higher classification accuracies. The results presented in [72] give a hint for this behavior since we showed that distortion correction introduces interpolation artifacts which have a negative impact on the classification performance. But we also showed that barrel-type distortions are problematic due to the distortions introduced farther away from the optical axis.

Since the LBP operator delivered promising results in case of the pit pattern images we investigated different LBP variants in a recent work [97]. In this work we proposed an adaptive Local Ternary Patterns (LTP) operator [94], an optimized quantization algorithm for contrast-based LBP (LBP/C) [81], and a wavelet based LBP operator. The results obtained in [97] indicated that the operators and extension proposed in this paper deliver classification accuracies which are highly competitive as compared to other LBP operators and previously developed methods.

Publications (sorted chronologically)

[44] S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vecsei. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 718–723, Salzburg, Austria, Sept. 2009

[98] A. Vécsei, T. Fuhrmann, M. Liedlgruber, L. Brunauer, H. Payer, and A. Uhl. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. *Computer Methods and Programs in Biomedicine*, 95:S68–S78, 2009

[27] M. Gschwandtner, M. Liedlgruber, A. Uhl, and A. Vécsei. Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, pages 1–6, Corfu, Greece, Nov. 2010

[97] A. Vécsei, G. Amann, S. Hegenbart, M. Liedlgruber, and A. Uhl. Automated marsh-like classification of celiac disease in children using an optimized local texture operator. *Computers in Biology and Medicine*, 41(6):313–325, June 2011

[72] M. Liedlgruber, A. Uhl, and A. Vécsei. Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'11)*, pages 1–6, Corfu, Greece, July 2011

3. Publications

This online version of the thesis provides hypertext links to the publishers' web sites where available as well as links to local copies of the respective PDF documents where permitted. The following copyright notices are reproduced here as required by the respective publishers.

[42] First published in the Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06) in 2006, published by IET.

[68] First published in the Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07) in 2008, published by WSEAS.

[44, 69, 32, 33, 36, 27, 72] © IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

[98, 97, 41] © Elsevier. The copyrights for these contributions are held by Elsevier. The original publications are available at http://www.sciencedirect.com.

[35] © Springer Verlag. The copyright for this contribution is held by Springer Verlag. The original publication is available at http://www.springerlink.com.

[70] Appeared in the book "Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain" in 2011, published by InTech.

[34] Appeared in the book "Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications" (2009), © IGI Global

[71] First published as technical report 2011-01 at the Department of Computer Sciences, University of Salzburg, Austria (2011), available from http://www.cosy.sbg.ac.at/research/tr.html

M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using wavelet texture features. In W. Sandham, D. Hamilton, and C. James, editors, *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06)*, pages 1–4, Glasgow, Scotland, UK, July 2006

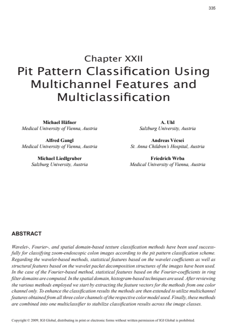

PIT PATTERN CLASSIFICATION OF ZOOM-ENDOSCOPIC COLON IMAGES USING WAVELET TEXTURE FEATURES

M. Liedlgruber and A. Uhl. Statistical and structural wavelet packet features for Pit pattern classification in zoom-endoscopic colon images. In P. Dondon, V. Mladenov, S. Impedovo, and S. Cepisca, editors, *Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07)*, pages 147–152, Arcachon, France, Oct. 2007.


Statistical and Structural Wavelet Packet Features for Pit Pattern Classification in Zoom-Endoscopic Colon Images

M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using multichannel features and multiclassification. In D. F. T.P. Exarchos, A. Papadopoulos, editor, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, pages 335–350. IGI Global, Hershey, PA, USA, 2009.

http://dx.doi.org/10.4018/978-1-60566-314-2


Chapter XXII
Pit Pattern Classification Using Multichannel Features and Multiclassification

M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Combining Gaussian Markov random fields with the discrete wavelet transform for endoscopic image classification. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'09)*, pages 177–182, Santorini, Greece, July 2009.


COMBINING GAUSSIAN MARKOV RANDOM FIELDS WITH THE DISCRETE WAVELET TRANSFORM FOR ENDOSCOPIC IMAGE CLASSIFICATION

M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using extended local binary patterns. In *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, pages 1–4, Larnaca, Cyprus, Nov. 2009.

http://dx.doi.org/10.1109/ITAB.2009.5394423


Pit Pattern Classification using Extended Local Binary Patterns

M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba

S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vecsei. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 718–723, Salzburg, Austria, Sept. 2009.


Impact of Duodenal Image Capturing Techniques and Duodenal Regions on the Performance of Automated Diagnosis of Celiac Disease

S.Hegenbart, R.Kwitt, M.Liedlgruber, A.Uhl
Department of Computer Sciences
Salzburg University, Austria

A.Vécsei
St. Anna Children's Hospital
Vienna, Austria

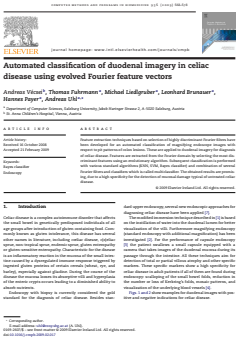
M. Liedlgruber and A. Uhl. Endoscopic image processing - an overview. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 707–712, Salzburg, Austria, Sept. 2009.


Endoscopic Image Processing - An Overview

M. Liedlgruber, A. Uhl
Department of Computer Sciences
University of Salzburg, Austria
mliedl@cosy.sbg.ac.at

A. Vécsei, T. Fuhrmann, M. Liedlgruber, L. Brunauer, H. Payer, and A. Uhl. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. *Computer Methods and Programs in Biomedicine*, 95:S68–S78, 2009.

http://dx.doi.org/10.1016/j.cmpb.2009.02.017


Automated classification of duodenal imagery in celiac disease using evolved Fourier feature vectors

Andreas Vécsei[1], Thomas Fuhrmann*, Michael Liedlgruber*, Leonhard Brunauer*, Hannes Payer*, Andreas Uhl*,*

[1] Department of Computer Sciences, Salzburg University, Jakob-Haringer-Strasse 2, A–5020 Salzburg, Austria
St. Anna Children's Hospital, Vienna, Austria

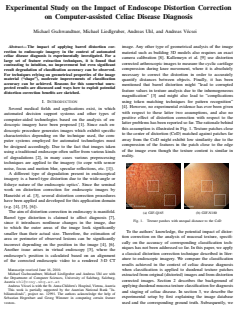
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Classification of endoscopic images using Delaunay triangulation-based edge features. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, volume 6112 of *Springer LNCS*, pages 131–140, Povoa de Varzim, Portugal, June 2010.

http://dx.doi.org/10.1007/978-3-642-13775-4_14

M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Endoscopic image classification using edge-based features. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, pages 2724–2727, Istanbul, Turkey, Aug. 2010.

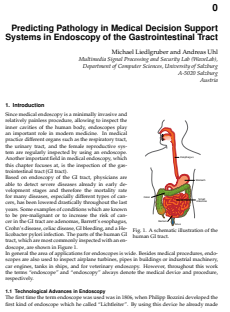
http://dx.doi.org/10.1109/ICPR.2010.667

M. Gschwandtner, M. Liedlgruber, A. Uhl, and A. Vécsei. Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, pages 1–6, Corfu, Greece, Nov. 2010.



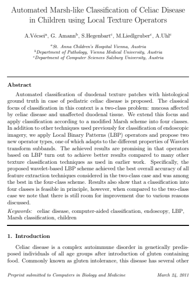
http://dx.doi.org/10.1109/ITAB.2010.5687708

M. Liedlgruber and A. Uhl. Predicting pathology in medical decision support systems in endoscopy of the gastrointestinal tract. In C. Jao, editor, *Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain*, pages 195–214. InTech, Rijeka, Croatia, 2011.



A. Vécsei, G. Amann, S. Hegenbart, M. Liedlgruber, and A. Uhl. Automated marsh-like classification of celiac disease in children using an optimized local texture operator. *Computers in Biology and Medicine*, 41(6):313–325, June 2011.

http://dx.doi.org/10.1016/j.compbiomed.2011.03.009


Automated Marsh-like Classification of Celiac Disease in Children using Local Texture Operators

A.Vécsei, G. Amann, S.Hegenbart, M.Liedlgruber, A.Uhl

M. Liedlgruber, A. Uhl, and A. Vécsei. Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'11)*, pages 1–6, Corfu, Greece, July 2011.

http://dx.doi.org/10.1109/ICDSP.2011.6004900


STATISTICAL ANALYSIS OF THE IMPACT OF DISTORTION (CORRECTION) ON AN AUTOMATED CLASSIFICATION OF CELIAC DISEASE

M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Medical Image Analysis*, 16(1):75–86, January 2012.

http://dx.doi.org/10.1016/j.media.2011.05.006


Color Treatment in Endoscopic Image Classification using Multi-scale Local Color Vector Patterns

M. Liedlgruber and A. Uhl. A summary of research targeted at computer-aided decision support in endoscopy of the gastrointestinal tract. Technical Report 2011-01, Department of Computer Sciences, University of Salzburg, Austria, 2011.

http://www.cosy.sbg.ac.at/research/tr.html


A Summary of Research Targeted at Computer-Aided Decision Support in Endoscopy of the Gastrointestinal Tract

4. Experiments and Discussion

Throughout the past years the image databases used in our work have been extended on a quite regular basis. This accounts to the pit pattern images as well as to the celiac disease images. As a consequence, our publications are based on different image databases which hardly allows a direct comparison of the classification results obtained by the different methods.

We therefore decided to repeat selected experiments on the most recent image databases. The respective set of methods which should be used for the new set of experiments was chosen based on two criteria: first, the methods used must have been originally reported in one of the publications listed in sections 2.2 or 2.3 to ensure a certain level of contribution from the author of this work. Second, we decided to only use methods which delivered promising results already in the respective earlier publications.

In the following we list and briefly describe the methods which fulfilled these two criteria and are therefore used for a set of new experiments.

WT-LDB [42] The Local Discriminant Basis algorithm (LDB) [86] has been developed in order to facilitate finding an optimal wavelet packet decomposition basis with respect to discrimination between different classes of images. We use the LDB algorithm to find an optimal basis into which all images are transformed to. Based on the resulting wavelet decompositions for each subband we use the variance of the coefficients contained within a subband as feature. Instead of using all subbands available for feature extraction we select a subset of S subbands which exhibit the highest discriminative power. The feature values computed from these subbands make up the feature vector.

When applied to color images, this operator is applied to each color channel separately and the final feature vector for an image is obtained by concatenating the feature vectors from the single channels.

WPC [42] This method applies the DWT to the input images. Based on the resulting wavelet decomposition structures, we compute the energy or the l-norm from the coefficients of a subband to obtain the feature for the respective subband. Similar to WT-LDB, this method also selects a subset of subbands S for feature extraction (chosen based on the l-norm computed from the coefficients within a subband). The feature values computed from these subbands make up the feature vector.

When applied to color images, this operator is applied to each color channel separately and the final feature vector for an image is obtained by concatenating the feature vectors from the single channels.

WT-BBC [68] The Best Basis Centroids method uses the Best-Basis algorithm [16] to find an optimal basis for each image in a training set and computes a centroid over all resulting Wavelet packet decomposition structures. After transforming all images into this basis, the most informative subset of S subbands is selected to compute a feature from all coefficients within each subband (energy, variance, or l-norm). The selection of the subset of subbands is based on a cost-function computed on all coefficients within a subband (log-energy, entropy, or l-norm). The feature values computed from the most informative subbands form the feature vector.

When applied to color images, this operator is applied to each color channel separately and the final feature vector for an image is obtained by concatenating the feature vectors from the single channels.

DELAUNAY [35] In this method the first step is to apply a block-based LBP operator [33] to an image, followed by making the transformation result rotation invariant. After performing a thresholding and applying a set of morphological operators, the resulting binary image is used to extract edges and compute the polygon centers for all objects found. After computing the Delaunay triangulation [7] based on these object centers, we construct histograms from the edge lengths of the Delaunay triangles.

To incorporate information from different color channels the edge-length histograms from different channels are concatenated to obtain the final histogram for an image.

This method has originally been developed with the aim of pit pattern classification of the colonic mucosa.

EDGEFEATURES [36] After applying the Canny Edge detector in order to detect edge-enclosed regions, different shape features and texture features based on the regions found are extracted. To find the best performing set of features a greedy forward feature selection is used.

In order to account for color information in the input images the feature selection considers features from all different channels.

Similar to the DELAUNAY-method this method has originally been developed with the aim of pit pattern classification of the colonic mucosa.

GMRF [32] In this method the different color channels of an input image are subject to a parameter estimation of a GMRF, using neighborhoods of Geman-type [21]. To obtain the final feature vector for classification the Markov parameters estimated along with the approximation error for each color channel are concatenated.

WT-GMRF [32] This method is an extension to the GMRF method. While the GMRF method operates in the spatial domain, the WT-GMRF method is based on wavelet-transformed color channels in order to be able to capture texture details at different resolutions. For this purpose each color channel of an input image is transformed using the DWT. Then the Markov parameters are estimated for each resulting subband (except for the approximation subband). In order to obtain the final feature vector for classification, the parameter vectors (again containing the approximation errors too) from all subbands and color channels are concatenated. The rather time-consuming feature selection used in [32] is neglected in this work.

WT-GMRF-CNH [32] In order to be able to better capture details from the different wavelet subbands (horizontal, vertical, and diagonal) we extended the WT-GMRF method by replacing the Geman-type neighborhoods by neighborhoods specifically tailored to the differently orientated subbands. The rather time-consuming feature selection used in [32] is neglected in this work.

JC-MB-LBP [33] Due to the noise-sensitive nature of the LBP operator we developed an operator which is less sensitive to noise by computing LBP numbers from averaged pixel blocks instead of single pixels. In addition, the resulting features also consider inter-channel relationships between different color channels by construction joint-histograms across multiple color channels.

LCVP [41] The LCVP operator is based on the idea of the JC-MB-LBP operator. But instead of computing the LBP transform for each color channel separately or computing joint-histograms, this operator treats an image as a color vector field. Based on suitable similarity measures between two color vectors, a compact histogram descriptor, incorporating all color information available, is computed for an input image. Similar to the idea of block averaging in the JC-MB-LBP operator the LCVP operator also uses a parameter which allows to investigate an image at different scales, which is the basis for the LCVP multi-scale operator.

While it is quite possible that some parameter combinations used throughout earlier work were not optimal we decided to stick to the parameters used in the publications where the respective methods are introduced. This way we are able to compare the methods as published in terms of the classification rates obtained on the new image databases.

4.1. Experimental Setup and Ground Truth Information

In this section we provide details about the experimental setup used for our new set of experiments and a detailed ground truth information for the pit pattern images as well as for the celiac disease images.

As already stated above, we decided to not again optimize the parameter configurations. This also accounts to an eventual pre-processing of the imagery prior to feature extraction and classification. Hence, some methods use pre-processing (e.g. Gaussian smoothing, Laplace sharpening [24], or an adaptive contrast enhancement [102]) while other methods rely on the unaltered images. Details on eventual pre-processing steps for a certain method can be gathered from the respective publication.

From our publications one also notices that throughout the years we employed different types of classifiers in order to evaluate our methods. The set of classifiers used consists of the k-NN classifier, the Bayes classifier, and Support Vector Machines (SVM) [19]. In the experiments presented in Section 4.2 we decided to use the same set of classifiers for each method as used in the original publication. However, because of the additional amount of time needed to carry out the experiments in case of the SVM classifier we decided to neglect this classifier completely.

Since in case of the pit pattern images the image database is quite limited in terms of the number of images available per class we are not able to assemble separate training and validation sets for this image database. While in case of the celiac disease images the number of available images would have allowed this we decided to employ cross-validation strategies in both cases in order to estimate the accuracy of the different methods. While we used the Leave-One-Out Cross-Validation (LOO-CV) protocol in most of our earlier work, in [41] the more meaningful and reliable Leave-One-Patient-Out Cross-Validation (LOPO-CV) protocol has been used. For the experiments presented in this thesis we carry out all experiments using both cross-validation protocol, which allows us to compare those. However, it must be noted that, at least in case of the pit pattern images, LOPO-CV is too restrictive, since it would be sufficient to exclude images of the same polyp from the training database in order to avoid overfitting. But since we do not have the information at hand on how many images have been taken from each polyp in the image database, we decided to go with the more restricted but safe option.

In order to be able to obtain meaningful numbers concerning the runtime performance of the different methods all timing tests have been carried out on a machine equipped with an Intel Core2Quad CPU at 2.83 GHz (single-threaded), running Linux.

Histology	Pit Pattern	6 classes [61]			3 classes [55]			2 classes		
		N_O	N_E	N_P	N_O	N_E	N_P	N_O	N_E	N_P
Normal	I	30	124	11	72	198	14	72	198	14
Hyperplasia	II	33	74	8						
Tubular adenoma	III-L	48	120	9	212	420	27	255	518	32
Tubulovillous adenoma		1	4							
Serrated adenoma	III-S	6	8	4						
Tubular adenoma		8	12							
Adenoma	IV	2	3	20						
Tubular adenoma		3	5							
Tubulovillous adenoma		144	268							
Adenocarcinoma	V	31	73	6	43	98	6			
Carcinoma		6	16							
Lymphoma		6	9							
	Total	**327**	**716**	**58**	**327**	**716**	**47**	**327**	**716**	**46**

Table 4.1.: The detailed ground truth information for the pit pattern image database used throughout our experiments.

4.1.1. Pit Pattern Images Ground Truth

The pit pattern image database used throughout our experiments is based on 327 endoscopic color images (either of size 624×533 pixels or 586×502 pixels) acquired between the years 2005 and 2009 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-colonoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. In order to acquire the images 40 patients underwent colonoscopy. To obtain a larger set of images we extracted subimages with a size of 256×256 pixels from the original images, which resulted in an extended image set containing 716 images in total.

Lesions found during colonoscopy have been examined after application of dye-spraying with indigocarmine, as routinely performed in colonoscopy. Biopsies or mucosal resection have been performed in order to get a histopathological diagnosis. Biopsies have been taken from type I, II, and type V lesions, as those lesions need not to be removed or cannot be removed endoscopically. Type III and IV lesions have been removed endoscopically. Out of all images from the extended set, histopathological classification resulted in 198 non-neoplastic and 518 neoplastic cases.

While in previously published work we focused on the 2-classes and 6-classes case only, we now also perform a 3-class classification according to [55], which groups the six different pit pattern types into normal lesions (pit pattern types I and II), non-invasive lesions (pit pattern types III-S, III-L, and IV), and invasive lesions (pit pattern type V). This classification scheme is of particular importance since normal mucosa needs not to be removed, non-invasive lesions must be removed endoscopically, and invasive lesions must not be removed endoscopically.

Table 4.1 shows the detailed ground truth information used for our experiments where $\mathbf{N_O}$, $\mathbf{N_E}$, $\mathbf{N_P}$ denote the number of original images, the number of images in the extended image set, and the number of patients in each class, respectively. Since different types of lesions may develop inside the colon of a single patient such a patient may appear in more than one class. Hence, the number of patients is slightly higher as compared to the total number of patients who underwent colonoscopy.

Modified Marsh type [80]	4 classes			2 classes		
	N_O	N_E	N_P	N_O	N_E	N_P
Marsh-0	260	306	131	260	306	131
Marsh-3A	55	95	11			
Marsh-3B	64	114	13	184	306	40
Marsh-3C	65	97	16			
Total	**444**	**612**	**171**	**444**	**612**	**171**

Table 4.2.: The detailed ground truth information for the celiac disease image database used throughout our experiments.

4.1.2. Celiac Disease Images Ground Truth

The celiac disease database used throughout this work is based on 444 endoscopic images taken during duodenoscopies at the St. Anna Children's Hospital using gastroscopes without magnification (two Olympus GIF-Q165 endoscopes with an image resolution of 768×576 pixels and one Olympus GIF-N180 endoscope with an image resolution of 528×522 pixels). In order to acquire the images 171 patients underwent duodenoscopy. To obtain a larger set of images and focus on image regions with a higher image quality, we extracted subimages with a size of 128×128 pixels from the original images, which resulted in an extended image set containing 612 images in total.

The main indications for endoscopy were the diagnostic evaluation of dyspeptic symptoms, positive celiac serology, anemia, malabsorption syndromes, inflammatory bowel disease, and gastrointestinal bleeding. Images have been recorded using the modified immersion technique, which is based on the instillation of water into the duodenal lumen for better visibility of the villi. The tip of the gastroscope is inserted into the water and images of interesting areas are taken.

In order to generate ground truth for the texture patches used in experimentation, the condition of the mucosal areas covered by the images was determined by histological examination of biopsies from the corresponding regions. Severity of villous atrophy was classified according to the modified Marsh classification. The detailed ground truth information used in our experiments is shown in Table 4.2.

4.2. Results and Discussion

4.2.1. Runtime Performance

In order to be able to properly compare the different methods developed in the past in terms of the runtime performance we fixed certain parameters across different methods to match the parameters best which on average lead to the best overall classification rates:

WPC, WT-LDB, and WT-BBC These methods have been configured to use either a wavelet decomposition depth of 3 levels (WT-LDB and WPC) or 4 (WT-BBC). Moreover, each of these methods is configured to return feature vectors containing 10 features per color channel.

WT-GMRF and WT-GMRF-CNH Each of these methods uses a wavelet decomposition depth of 2 in order to ensure a stable parameter estimation in case of the celiac disease images (which, compared to the pit pattern images, are smaller). The neighborhood order was set to 1.

GMRF All tests carried out with this method use Geman neighborhoods of order 4.

JC-MB-LBP, LCVP In accordance to previous experiments the block width for the block averaging has been fixed to 5 pixels. The multi-scale experiments (denoted by MR) use the filter kernel widths 3, 5, and 7. While for these methods we carried out experiments in the RGB color space as well as in the LAB color space, the performance measurements have been obtained for the RGB experiments only. Since color space conversions are already carried out as pre-processing steps there is no impact on the runtime performance.

DELAUNAY The number of bins for the edge length histograms has been fixed to 101 and the k-NN classifier used considers 13 neighbors.

EDGEFEATURES The number of features used has been fixed to 7 (we always use the same set of features) and the k-NN classifier used considers 5 neighbors.

Except for the DELAUNAY method and the EDGEFEATURES method the experiments based on the k-NN classifier have always been carried out with $k = 9$.

Since the number of classes has no measurable impact on the runtime measurements the timing tests have always been carried out for the 2-classes cases only. In addition the measurements have always been performed for the LOO-CV protocol since the choice for the cross-validation protocol has no influence on the timing measurements.

In order to compare the computational demand between different methods we measured the time needed to extract the respective features from an image ($\mathbf{T_O}$). In addition we measured the time needed for the classification ($\mathbf{T_C}$). Figures 4.1 and 4.2 show the results of these measurements.

A detailed presentation of the timing measurements carried out is given in Table 4.3, where we also present the total time needed by a method for a complete classification of a single image ($\mathbf{T_T}$). All these values have been obtained by extracting and classifying all images contained within the respective image databases and dividing the time needed by the number of images contained in each image database. This way the measurements are assumed to be way more stable as compared to measuring the time needed for one image only.

As we notice from Figures 4.1 and 4.2 JC-MB-LBP is clearly the slowest method in terms of the classification. This can be attributed to the fact that this method is based on 2D histograms which considerably slows down the histogram comparison used. In addition we see that the EDGEFEATURES method is faster in terms of the classification, this method is clearly the slowest method in terms of the time needed to extract features which is due to the rather high amount of different shape-based features extracted in this method. While the DELAUNAY method needs only about a fifth of the time needed to extract features in case of EDGEFEATURES this method nevertheless belongs to the most time-consuming methods in terms of the feature extraction. This is mainly due to the need to apply the Canny edge detector and the Delaunay triangulation.

The remaining methods (wavelet-based methods, GMRF-based methods, and the LCVP methods) roughly deliver an equal performance. An exception are the multi-scale extensions to LCVP (LCVP MR $A^{(1)}$ and LCVP MR $A^{(2)}$) which are slower as compared to their single scale counterparts. However, this is quite logical since the feature extraction is carried out for three different scales in case of the multi-scale experiments.

In order to be able to compare the time measurements among different methods more easily we provide the rough indicator **SF** in Table 4.3. This value is a multiplicative factor denoting the total time consumed for a complete classification of an image as compared to the fastest method (which is LCVP MR $A^{(2)}$ in case of both image databases). As we notice, most methods are up

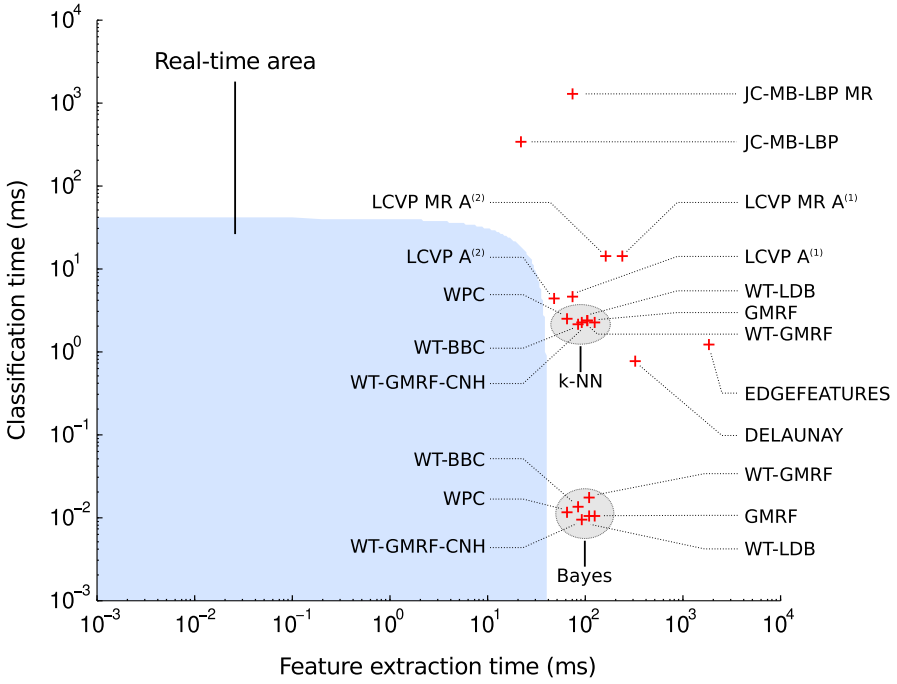


Figure 4.1.: Performance measurement results for the pit pattern images (256×256 pixels, RGB, 2 classes)

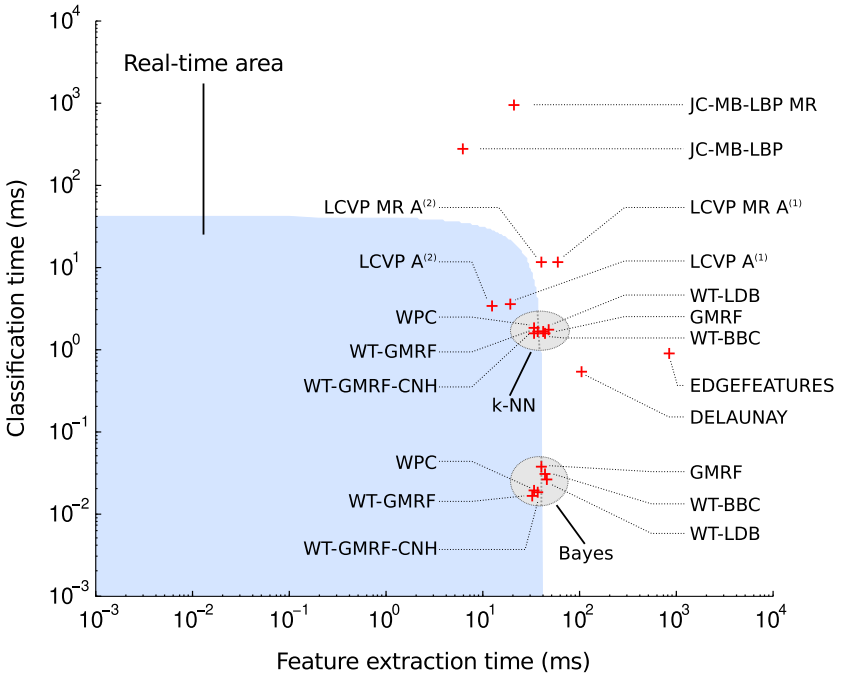


Figure 4.2.: Performance measurement results for the celiac disease images (128×128 pixels, RGB, 2 classes)

Method	Pit Pattern				Celiac Disease				Q
	$\mathbf{T_O}$	$\mathbf{T_C}$	$\mathbf{T_T}$	**SF**	$\mathbf{T_O}$	$\mathbf{T_C}$	$\mathbf{T_T}$	**SF**	
WT-LDB k-NN	106	2	108	2.1	47	2	49	3.1	2.2
WT-LDB Bayes	107	< 1	107	2.0	46	< 1	46	2.9	2.3
WPC k-NN	65	2	67	1.3	34	2	36	2.3	1.9
WPC Bayes	66	< 1	66	1.3	34	< 1	34	2.2	1.9
WT-BBC k-NN	83	2	85	1.6	44	2	45	2.9	1.9
WT-BBC Bayes	84	< 1	84	1.6	43	< 1	43	2.8	1.9
DELAUNAY	325	1	326	6.2	102	1	103	6.6	3.2
EDGEFEATURES	1768	1	1769	33.8	839	1	840	53.9	2.1
JC-MB-LBP	22	336	358	6.8	6	277	283	18.2	1.3
JC-MB-LBP MR	75	1289	1363	26.0	21	920	941	60.4	1.4
LCVP A$^{(1)}$	74	4	79	1.5	19	4	22	1.4	3.5
LCVP A$^{(2)}$	48	4	52	1.0	12	3	16	1.0	3.4
LCVP MR A$^{(1)}$	236	14	250	4.8	60	12	71	4.6	3.5
LCVP MR A$^{(2)}$	157	14	171	3.3	40	12	52	3.3	3.3
GMRF k-NN	124	2	127	2.4	42	2	43	2.8	2.9
GMRF Bayes	125	< 1	125	2.4	40	< 1	40	2.6	3.1
WT-GMRF k-NN	106	2	108	2.1	36	2	38	2.4	2.8
WT-GMRF Bayes	106	< 1	106	2.0	37	< 1	37	2.4	2.9
WT-GMRF-CNH k-NN	92	2	94	1.8	33	2	35	2.2	2.7
WT-GMRF-CNH Bayes	92	< 1	92	1.8	33	< 1	33	2.1	2.8

Table 4.3.: Comparison of time measurements given in milliseconds (feature extraction time T_O, classification time T_C, and total time T_T) for the different methods evaluated on two different image sets (pit pattern images – 256 $\times$ 256 pixels and celiac disease images – 128 $\times$ 128 pixels).

to approximately 5 times slower as compared to the fastest method. But, as already indicated earlier, in case of JC-MB-LBP, DELAUNAY, and EDGEFEATURES **SF** is quite high.

In Figure 4.3 we compare the performance measurements for the different methods between the pit pattern images and the celiac disease images. We notice that the methods mostly perform significantly faster in case of the celiac disease images, which is no surprise considering the different image dimensions. This difference gets also obvious when looking at the values for **Q** in Table 4.3. Based on the values of $\mathbf{T_O}$, this multiplicative value roughly indicates how much faster the methods perform in case of the celiac disease images (e.g. $\mathbf{Q} = 3.1$ means that the methods are about 3.1 times faster in case of celiac disease images as compared to the pit pattern images). The values for **Q** show that the methods are at least 1.3 times faster when using images of size 128 $\times$ 128 pixels instead of 256 $\times$ 256 pixels. Since, as already discussed above, in case of JC-MB-LBP the share of the feature extraction time is rather small it is clear that this is also the method which is affected the least by the image dimensions (in case of multi-scale the picture is rather similar). Due to the compact and efficient nature of LCVP this method shows the highest dependence on the image resolution.

From Figure 4.3 we also immediately see that the degree of contribution to the total image classification time differs significantly between feature extraction and classification. While for most methods the share of time needed for classification can be neglected, the picture is quite different for JC-MB-LBP. For this method the time needed for classification is about 16 times higher as compared to the feature extraction time (in the single scale case as well as in the multi-scale case). From Table 4.3 we can also easily see that, in case of the Bayes classifier, the time needed for the classification of an image ($\mathbf{T_C}$) is always considerably lower.

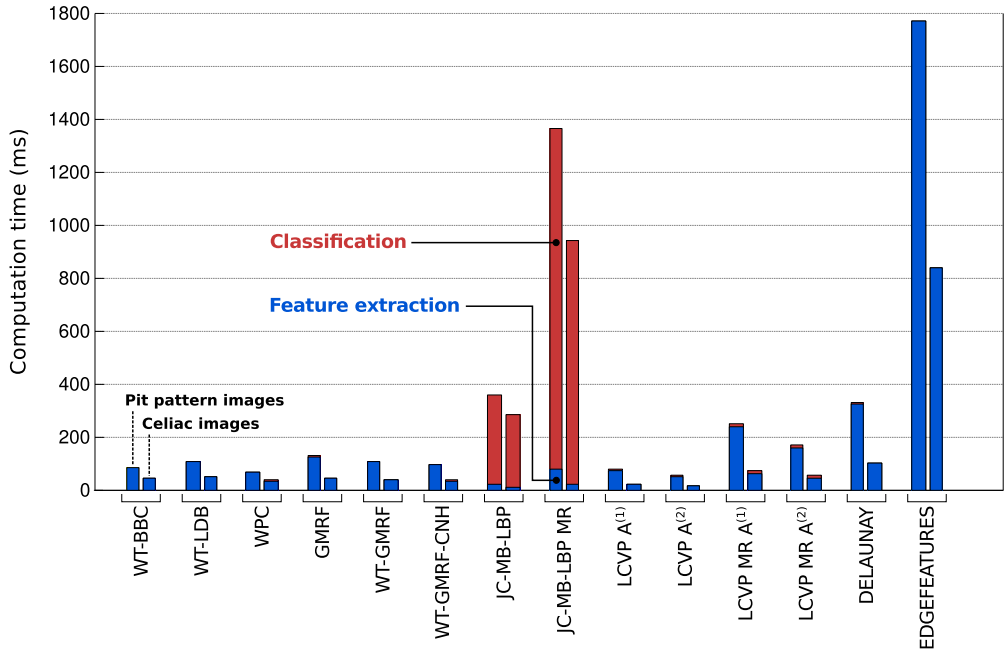


Figure 4.3.: Comparison of performance measurements between the pit pattern images (256×256) and the celiac disease images (128×128).

If we assume an endoscopy video to be captured at a frame rate of 25 frames per second a real-time application demands processing times of at most 40 milliseconds for a single frame. In order to be able to assess which of the different methods used are capable of performing a real-time diagnosis we sketched the respective real-time area in figures 4.1 and 4.2. As we quickly notice form Figure 4.1, in case of the pit pattern images there is no method which at the moment is suited for a real-time diagnosis, although LCVP MR $^{(2)}$ gets really close to meet the respective runtime constraints. While the overall picture looks quite similar in case of the celiac disease images all methods are shifted towards the real-time area and quite a few methods already meet the real-time constraints, as we notice from Figure 4.2. But since the celiac disease images are only a quarter in size as compared to the pit pattern images, this is not surprising at all. Furthermore it must be noted that by optimizing the respective implementations or using a faster machine for the experiments even more methods would meet the real-time constraints, also in case of the pit pattern images.

4.2.2. Classification Results

In the following we present the classification results achieved by the different methods on our image databases. The configurations used for the different methods are the ones which lead to the highest overall classification results. To find these values we conducted a large variety of differently configured tests, evaluating a large pool of parameter configurations.

In order to be able to make an assessment about whether the classification outcome of two methods is statistically significant different we employ McNemar's test [20]. For two methods M_1 and M_2 this test statistic keeps track of the number of images which are misclassified by method M_1 but classified correctly by method M_2 (denoted by n_{01}) and vice versa (denoted

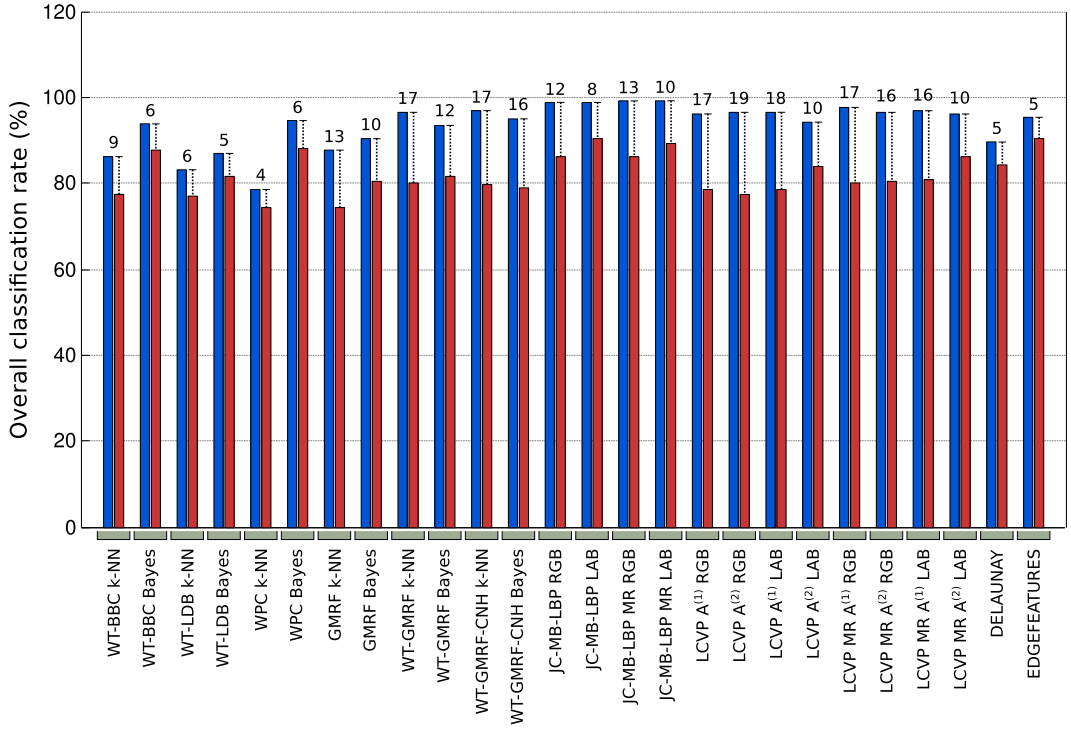


Figure 4.4.: Comparison of the overall classification rates for the pit pattern images where the blue and red bars denote the LOO-CV results and LOPO-CV results respectively (2 classes).

by n_{10}). The test statistic, which is approximately Chi Square distributed (with one degree of freedom), is then computed as

$$T = \frac{(|n_{01} - n_{10}| - 0.5)^2}{n_{01} + n_{10}}. \tag{4.1}$$

From T the p-value can be computed as

$$p = 1 - F_{\chi_1^2}(T) \tag{4.2}$$

where $F_{\chi_1^2}$ denotes the cumulative distribution function of the Chi Square distribution with one degree of freedom. The null-hypothesis H_0 for McNemar's test is that the outcomes of M_1 and M_2 lead to equal error rates. Given a fixed significance level α, there is evidence that the methods M_1 and M_2 produce significantly different results if $p < \alpha$. As a consequence we can reject the null-hypothesis H_0. For the discussion in the next sections we chose a significance level of $\alpha = 0.05$. This implies that, if M_1 and M_2 are significantly different, there is a confidence level of 95% that the differences between the outcomes of the methods are not caused by random variation.

Pit Pattern Images

Figures 4.4, 4.5, and 4.6 show the overall classification rates obtained with the different methods on the pit pattern image database. In each of this figures the blue bars denote the results

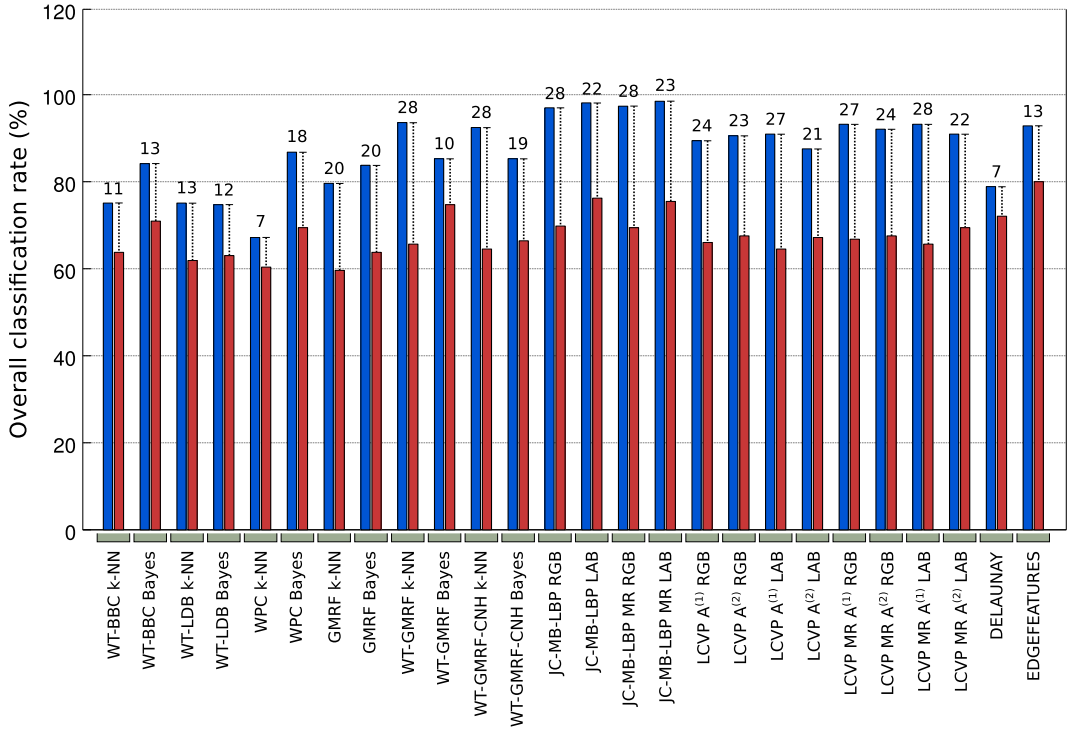


Figure 4.5.: Comparison of the overall classification rates for the pit pattern images where the blue and red bars denote the LOO-CV results and LOPO-CV results respectively (3 classes).

obtained with LOO-CV, while the the red bars denote the results obtained using LOPO-CV. The numbers on top of each pair of bars indicate the result drop when switching from LOO-CV to LOPO-CV. In addition we provide an indicator next to the method names, which – if filled gray – shows that an eventual result drop is statistically significant according to McNemar's test.

The detailed classification rates for the 2-classes case, the 3-classes case, and the 6-classes case can be found in tables 4.4, 4.5, and 4.6, respectively. In these tables we provide the detailed classification rates for the single classes in case of LOO-CV. In addition we provide the result drop when using LOPO-CV instead of LOO-CV for each result in brackets.

The column $\mathbf{S_C V}$ indicates whether an eventual result difference between the observed overall rates of LOO-CV and LOPO-CV is statistically significant according to McNemar's test or not (statistical significance is shown by a ✓). If the differences between the overall rates are statistically significant we also provide an indicator which shows if in case of LOPO-CV the results dropped (shown by a "$-$") or have been improved (shown by a "$+$").

From figures 4.4, 4.5, and 4.6 we notice that in case of LOO-CV the best overall classification rates have always been achieved by the JC-MB-LBP method, followed by the LCVP method, and EDGEFEATURES. In case of LOPO-CV the picture is rather similar. Again the best overall classification rates are obtained by one of these methods. From these figures we also notice that, compared to the other methods, the DELAUNAY method constantly performs worse the higher the number of classes used gets. This can be explained by the fact that, as also mentioned in [35], this method has originally been developed with a 2-class classification based on different the pit

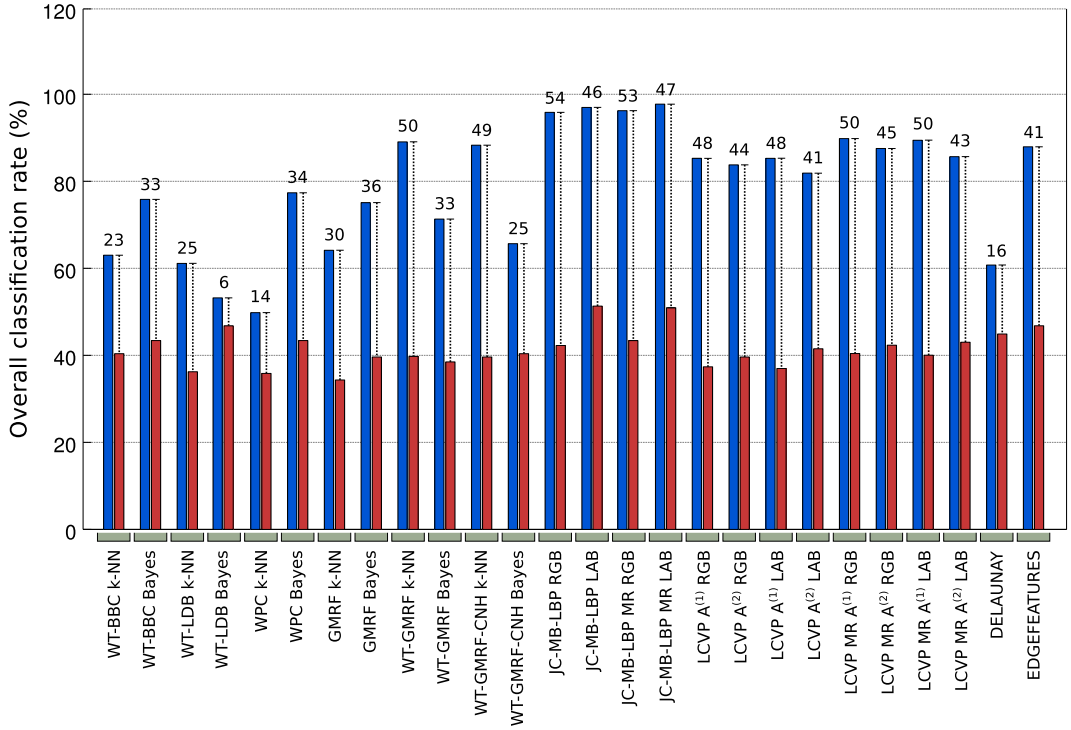


Figure 4.6.: Comparison of the overall classification rates for the pit pattern images where the blue and red bars denote the LOO-CV results and LOPO-CV results respectively (6 classes).

densities within the two classes in mind. In case of higher class counts, however, the different classes can not be distinguished that easy by such a feature since the difference in pit density is only clearly noticeable between non-neoplastic and neoplastic images. Within each of these two classes the densities are rather similar, hence, a finer classification scheme is problematic.

Nevertheless, as we notice from figures 4.4, 4.5, and 4.6 as well as from tables 4.4, 4.5, and 4.6, the overall classification accuracies always drop. This is especially noticeable for higher class counts. In addition, in case of the pit pattern images the result differences are always statistically significant. While in the 2-classes case the results drop by approximately 4% to 20 %, the decrease is rather huge in case of the 3-classes and 6-classes experiments with a loss of approximately 7% to 28% and 6% to 54%, respectively. One explanation for this behavior is that in order to obtain our extended image database we quite often had to cut out more than one subimage from one original endoscopic image (which shows one certain polyp from one specific patient). This however, in turn lead to an overfitting in case of LOO-CV since the training set very often contained an image which exhibited a high similarity to the one classified. In addition, as already pointed out in Section 4.1, LOPO-CV is very restrictive in case of the pit pattern images. Since each patient may be present in more than one class, it quite often happens, that leaving out images from a certain patient from the training set also results in the unnecessary reduction of training sets for other image classes. As a consequence it is quite likely that the training set for certain classes often is simply too small in order to allow a proper training and classification.

Method	Non-Neoplastic	Neoplastic	Total	S_{CV}
WT-LDB k-NN	68.7 (-24.7)	88.4 (1.4)	83.0 (-5.9)	✓ (−)
WT-LDB Bayes	71.7 (-37.4)	92.9 (6.9)	87.0 (-5.3)	✓ (−)
WT-BBC k-NN	65.2 (-22.2)	94.2 (-3.5)	86.2 (-8.7)	✓ (−)
WT-BBC Bayes	90.9 (-18.2)	95.0 (-1.4)	93.9 (-6.0)	✓ (−)
WPC k-NN	37.9 (-15.2)	94.0 (-0.2)	78.5 (-4.3)	✓ (−)
WPC Bayes	91.9 (-12.6)	95.8 (-4.1)	94.7 (-6.4)	✓ (−)
DELAUNAY	67.2 (-12.1)	98.1 (-2.7)	89.5 (-5.3)	✓ (−)
EDGEFEATURES	88.4 (-12.1)	98.1 (-2.5)	95.4 (-5.2)	✓ (−)
JC-MB-LBP RGB	96.0 (-31.3)	99.6 (-5.2)	98.6 (-12.4)	✓ (−)
JC-MB-LBP LAB	98.0 (-22.7)	99.2 (-2.9)	98.9 (-8.4)	✓ (−)
JC-MB-LBP MR RGB	98.5 (-34.3)	99.4 (-4.8)	99.2 (-13.0)	✓ (−)
JC-MB-LBP MR LAB	98.5 (-15.2)	99.4 (-8.1)	99.2 (-10.1)	✓ (−)
LCVP $A^{(1)}$ RGB	89.4 (-45.5)	98.6 (-6.8)	96.1 (-17.5)	✓ (−)
LCVP $A^{(2)}$ RGB	91.4 (-60.1)	98.6 (-3.5)	96.6 (-19.1)	✓ (−)
LCVP $A^{(1)}$ LAB	89.4 (-45.5)	99.0 (-7.1)	96.4 (-17.7)	✓ (−)
LCVP $A^{(2)}$ LAB	86.4 (-33.8)	97.1 (-1.2)	94.1 (-10.2)	✓ (−)
LCVP MR $A^{(1)}$ RGB	93.4 (-51.0)	99.0 (-4.4)	97.5 (-17.3)	✓ (−)
LCVP MR $A^{(2)}$ RGB	92.4 (-36.4)	98.3 (-8.3)	96.6 (-16.1)	✓ (−)
LCVP MR $A^{(1)}$ LAB	91.9 (-43.4)	99.0 (-5.6)	97.1 (-16.1)	✓ (−)
LCVP MR $A^{(2)}$ LAB	89.9 (-32.3)	98.6 (-1.7)	96.2 (-10.2)	✓ (−)
GMRF k-NN	81.8 (-44.9)	90.0 (-1.2)	87.7 (-13.3)	✓ (−)
GMRF Bayes	82.3 (-2.0)	93.6 (-13.1)	90.5 (-10.1)	✓ (−)
WT-GMRF k-NN	93.4 (-48.0)	97.9 (-4.8)	96.6 (-16.8)	✓ (−)
WT-GMRF Bayes	86.4 (-37.9)	95.9 (-1.9)	93.3 (-11.9)	✓ (−)
WT-GMRF-CNH k-NN	93.4 (-53.0)	98.1 (-3.5)	96.8 (-17.2)	✓ (−)
WT-GMRF-CNH Bayes	85.9 (-48.0)	98.5 (-3.7)	95.0 (-15.9)	✓ (−)

Table 4.4.: The detailed classification results obtained with the different methods evaluated on the pit pattern images (2 classes).

A more detailed look at the results for the different classes in tables 4.4, 4.5, and 4.6 reveals that for each class count there is one specific class, which mostly exhibits the highest result drops when using LOPO-CV instead of LOO-CV. While in case of the 2-classes experiments this is the class containing the non-neoplastic images, in the 3-classes case this is the class showing invasive lesions. For the 6-classes case the most significant result drops can be observed for pit pattern type III-S. An explanation for this behaviour is the fact that in all these cases the affected image classes contain a significantly lower number of images as compared to the other classes. This especially applies to class III-S in case of the 6-classes case. As a consequence, when using LOPO-CV, a rather high number of images per patient gets removed from the training set which has a negative impact on the learning ability of the classifier used.

When comparing the result of JC-MB-LBP as well as the result for LCVP in terms of the color space used, we can see from tables 4.4, 4.5, and 4.6 that the observed differences in terms of the overall classification accuracies are mostly only marginal. Apart from that, there is no obvious tendency noticeable whether these methods perform clearly better in either the RGB or the LAB color space. But the drops in case of LOPO-CV are most times lower in case of the LAB experiments compared to the result drops in case of the RGB experiments.

In figures 4.7 to 4.12 we provide some sort of method ranking in order to give a different view on the overall classification rates achieved by the different methods when applied to the pit pattern image database. These figures show the various methods sorted by the overall ac-

Method	Normal	Non-Invasive	Invasive	Total	S_{CV}
WT-LDB k-NN	70.7 (-19.7)	81.9 (-0.2)	54.1 (-54.1)	75.0 (-13.0)	✓ (−)
WT-LDB Bayes	67.7 (-12.6)	91.2 (-12.6)	17.3 (-5.1)	74.6 (-11.6)	✓ (−)
WT-BBC k-NN	63.1 (-14.1)	89.0 (-3.3)	39.8 (-39.8)	75.1 (-11.3)	✓ (−)
WT-BBC Bayes	89.9 (-19.2)	87.9 (0.0)	58.2 (-58.2)	84.4 (-13.3)	✓ (−)
WPC k-NN	44.9 (-19.2)	89.8 (1.0)	15.3 (-15.3)	67.2 (-6.8)	✓ (−)
WPC Bayes	93.4 (-13.6)	91.7 (-11.0)	54.1 (-54.1)	87.0 (-17.6)	✓ (−)
DELAUNAY	72.7 (-4.0)	90.7 (-0.7)	41.8 (-39.8)	79.1 (-7.0)	✓ (−)
EDGEFEATURES	90.4 (-20.2)	96.2 (-4.3)	84.7 (-35.7)	93.0 (-13.0)	✓ (−)
JC-MB-LBP RGB	96.0 (-29.3)	99.3 (-16.4)	90.8 (-71.4)	97.2 (-27.5)	✓ (−)
JC-MB-LBP LAB	98.0 (-22.2)	98.8 (-10.0)	95.9 (-72.4)	98.2 (-21.9)	✓ (−)
JC-MB-LBP MR RGB	96.0 (-34.3)	99.5 (-10.7)	92.9 (-90.8)	97.6 (-28.2)	✓ (−)
JC-MB-LBP MR LAB	98.5 (-15.2)	98.8 (-19.5)	96.9 (-54.1)	98.5 (-23.0)	✓ (−)
LCVP $A^{(1)}$ RGB	89.4 (-39.4)	93.1 (-6.4)	74.5 (-65.3)	89.5 (-23.6)	✓ (−)
LCVP $A^{(2)}$ RGB	91.4 (-50.0)	94.8 (-1.4)	72.4 (-63.3)	90.8 (-23.3)	✓ (−)
LCVP $A^{(1)}$ LAB	89.4 (-41.4)	94.3 (-9.0)	81.6 (-71.4)	91.2 (-26.5)	✓ (−)
LCVP $A^{(2)}$ LAB	86.4 (-37.4)	91.9 (-6.2)	72.4 (-49.0)	87.7 (-20.7)	✓ (−)
LCVP MR $A^{(1)}$ RGB	93.4 (-39.9)	95.2 (-9.5)	85.7 (-73.5)	93.4 (-26.7)	✓ (−)
LCVP MR $A^{(2)}$ RGB	92.4 (-46.0)	95.0 (-6.7)	79.6 (-57.1)	92.2 (-24.4)	✓ (−)
LCVP MR $A^{(1)}$ LAB	91.9 (-38.4)	96.0 (-13.3)	85.7 (-67.3)	93.4 (-27.7)	✓ (−)
LCVP MR $A^{(2)}$ LAB	89.9 (-30.3)	95.0 (-14.0)	76.5 (-35.7)	91.1 (-21.5)	✓ (−)
GMRF k-NN	81.8 (-42.4)	81.0 (-0.2)	69.4 (-59.2)	79.6 (-20.0)	✓ (−)
GMRF Bayes	90.4 (-23.2)	80.7 (-8.6)	83.7 (-62.2)	83.8 (-20.0)	✓ (−)
WT-GMRF k-NN	93.4 (-47.0)	95.5 (-10.0)	87.8 (-67.3)	93.9 (-28.1)	✓ (−)
WT-GMRF Bayes	84.8 (-10.6)	92.9 (-0.2)	54.1 (-54.1)	85.3 (-10.5)	✓ (−)
WT-GMRF-CNH k-NN	93.4 (-51.5)	95.0 (-8.3)	80.6 (-66.3)	92.6 (-28.2)	✓ (−)
WT-GMRF-CNH Bayes	87.4 (-43.4)	97.9 (-5.2)	28.6 (-28.6)	85.5 (-19.0)	✓ (−)

Table 4.5.: The detailed classification results obtained with the different methods evaluated on the pit pattern images (3 classes).

curacies reached (sorted ascending from left to right). In addition, we evaluated the statistical significance of eventual differences between each method and the worst performing (left-most method) and best performing method (right-most method) in terms of the overall classification accuracies for these figures. This resulted in three differently colored regions into which each method evaluated belongs to, according to the following rules:

Red area
Methods falling into this region achieve an overall classification accuracy which is significantly lower as compared to the best performing method. In addition these methods differ insignificantly only from the worst performing method.

Orange area
Methods falling into this category differ significantly from the best performing method as well as from the worst performing method in terms of the overall classification accuracies achieved.

Green area
Methods falling into this region achieve an overall classification accuracy which are not significantly different as compared to the best performing method. In addition these methods differ significantly from the worst performing method.

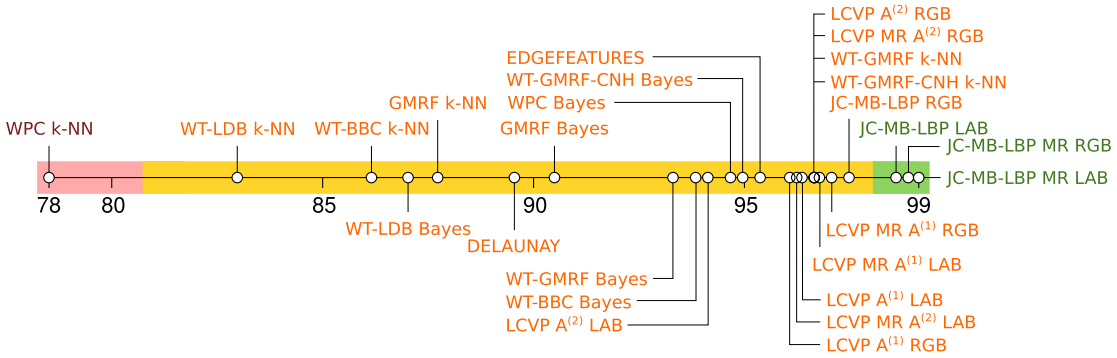


Figure 4.7.: Method ranking for the methods when applied to the pit pattern image database (2 classes, LOO-CV).

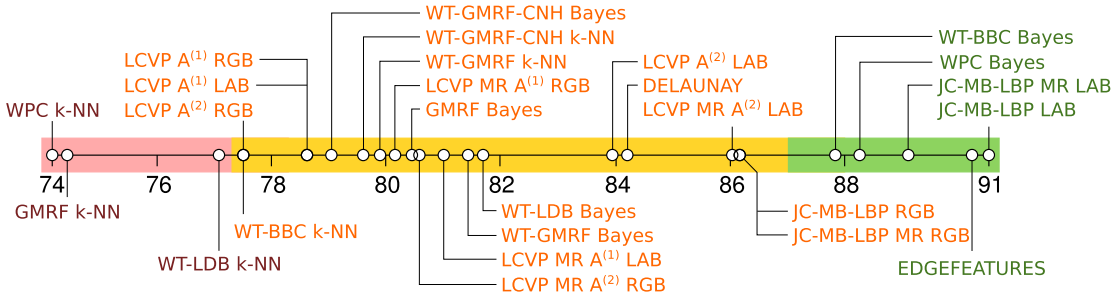


Figure 4.8.: Method ranking for the methods when applied to the pit pattern image database (2 classes, LOPO-CV).

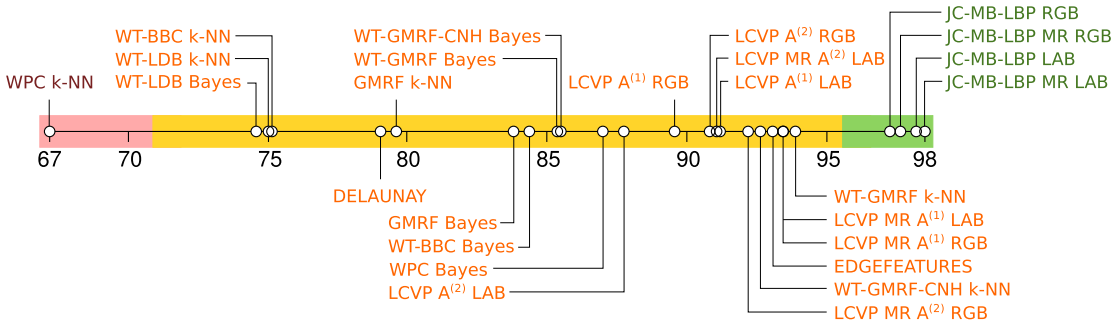


Figure 4.9.: Method ranking for the methods when applied to the pit pattern image database (3 classes, LOO-CV).

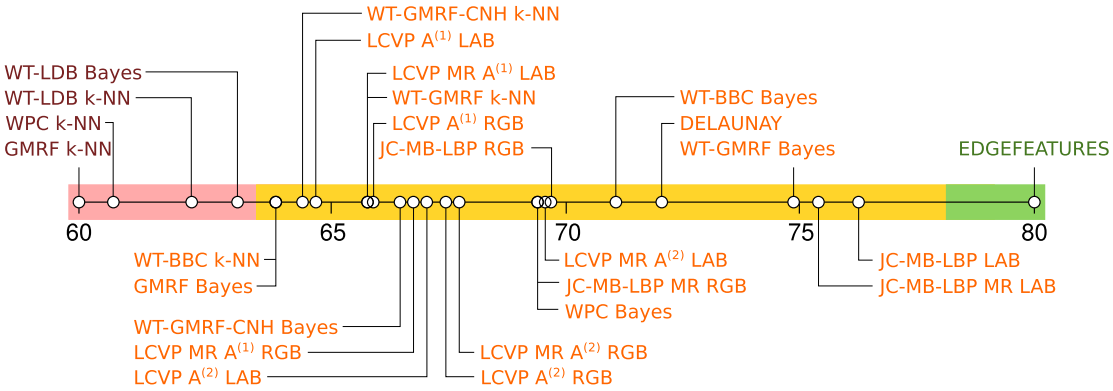


Figure 4.10.: Method ranking for the methods when applied to the pit pattern image database (3 classes, LOPO-CV).

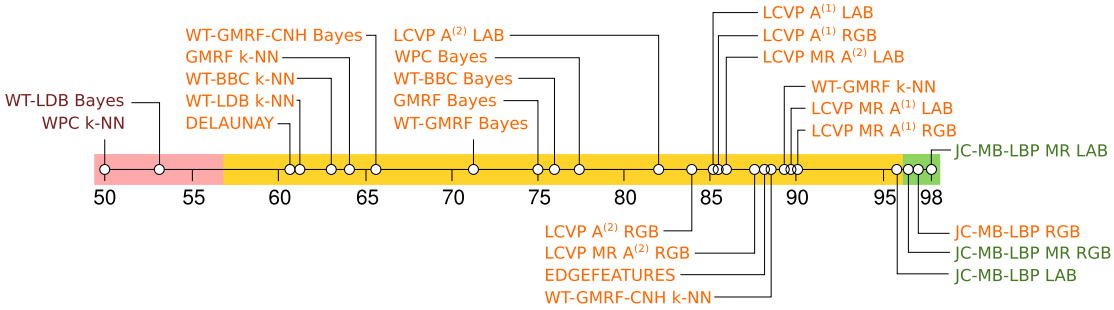


Figure 4.11.: Method ranking for the methods when applied to the pit pattern image database (6 classes, LOO-CV).

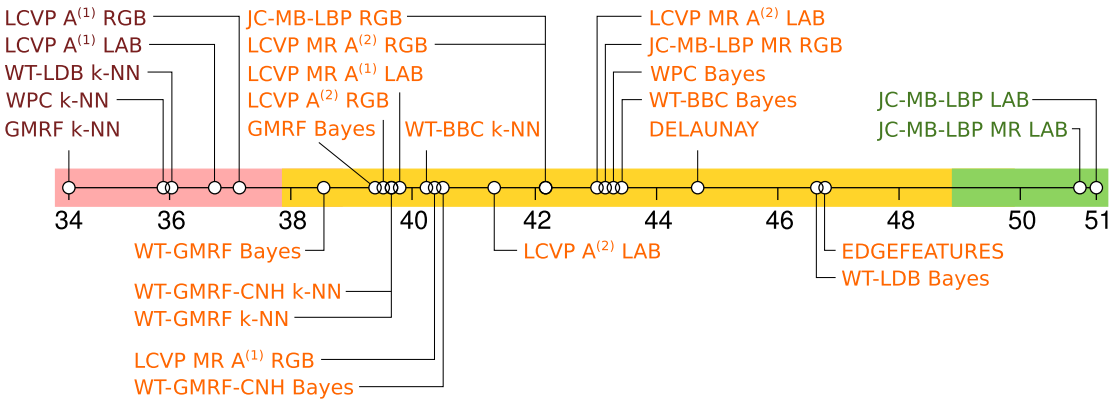


Figure 4.12.: Method ranking for the methods when applied to the pit pattern image database (6 classes, LOPO-CV).

It must be noted, that the borders between the differently colored regions have been introduced artificially. This means that the borders have been computed as the midpoint between the classification rate of the right-most method from the left region and the left-most method of the right region (for the red-orange and orange-green region pairs).

Hence, while for example Figure 4.15 implies that a method having an overall classification accuracy of 80% would automatically fall into the red region, this is not necessarily the case, since the membership to a region is only dependent on the statistical significance between a method and the best and worst performing methods. Thus, if the outcome of such a method is significantly different from WPC (k-NN), the border between the red and orange region would be shifted to the left, to lie in the middle between the new method and WPC (k-NN). If the outcome of such a method is not significantly different from WPC (k-NN), the border would be shifted to the right to lie between the new method and WT-LDB (k-NN).

The figures 4.7 to 4.12 clearly show, that for the pit pattern images most of the methods belong to the orange region in case of the different class counts as well as in the case of the different cross-validation protocols, which indicates that most methods perform significantly worse as compared to the best performing method. In addition we also notice that the methods get shifted to the left when using LOPO-CV. Hence, while most methods perform significantly better when compared to the worst method, the shift clearly shows that the pit pattern experiments tend to yield worse results when using LOPO-CV.

We also notice that the green regions are dominated by the JC-MB-LBP method, which underpins the superiority of this method. Only in case of the 3-classes case and LOPO-CV (see Figure 4.10) the EDGEFEATURES method delivers significantly higher results as compared to all other methods. One reason for this behavior is the fact that the EDGEFEATURES method delivers competitive results in case of LOO-CV already. When using LOPO-CV this method is able to better adapt to the missing training data due to the feature optimization employed. While other methods exhibit a similar or even lower result drop when switching from LOO-CV to LOPO-CV (e.g. WT-BBC Bayes and WPC k-NN) those methods perform rather poor in case of LOO-CV already.

From these figures we also notice that the ordering of the methods is similar, no matter whether we carry out a 2-class, 3-class, or 6-class classification. But when comparing the ranking plots for LOO-CV and LOPO-CV we notice that the ordering of the methods changes. This indicates that there are indeed methods in our method set which profited from the overfitting in case of LOO-CV and, hence, perform worse in case of LOPO-CV as compared to other methods (e.g. LCVP MR $A^{(1)}$ or WT-GMRF-CNH k-NN). Some methods prove to be rather stable as compared to other methods (e.g. JC-MB-LBP MR or WT-LDB k-NN). But there are also methods which seem to suffer from the overfitting and thus deliver a higher classification accuracy as compared to the other methods (e.g. WT-BBC Bayes or the DELAUNAY method).

Method	I	II	III-L	III-S	IV	V	Total	S_{CV}
WT-LDB k-NN	61.3 (-16.9)	50.0 (-21.6)	65.3 (-56.5)	50.0 (-50.0)	65.9 (-4.3)	53.1 (-52.0)	61.2 (-25.1)	✓ (−)
WT-LDB Bayes	64.5 (-64.5)	47.3 (-47.3)	71.0 (-17.7)	10.0 (-10.0)	47.8 (49.3)	43.9 (-43.9)	53.1 (-6.4)	✓ (−)
WT-BBC k-NN	65.3 (-24.2)	50.0 (-32.4)	66.1 (-58.1)	40.0 (-40.0)	69.9 (3.3)	51.0 (-38.8)	63.0 (-22.8)	✓ (−)
WT-BBC Bayes	83.9 (-29.8)	73.0 (-70.3)	75.8 (-61.3)	0.0 (0.0)	81.5 (-1.4)	68.4 (-65.3)	76.0 (-32.5)	✓ (−)
WPC k-NN	41.9 (-38.7)	32.4 (-12.2)	46.0 (-33.1)	0.0 (0.0)	74.6 (5.8)	18.4 (-18.4)	49.9 (-14.0)	✓ (−)
WPC Bayes	89.5 (-23.4)	47.3 (-39.2)	72.6 (-56.5)	0.0 (0.0)	94.2 (-23.9)	59.2 (-51.0)	77.4 (-34.1)	✓ (−)
DELAUNAY	49.2 (-28.2)	64.9 (-16.2)	33.1 (-32.3)	0.0 (0.0)	87.0 (0.7)	44.9 (-29.6)	60.6 (-15.9)	✓ (−)
EDGEFEATURES	87.9 (-25.8)	89.2 (-70.3)	91.1 (-81.5)	85.0 (-85.0)	86.6 (-4.7)	88.8 (-82.7)	88.1 (-41.3)	✓ (−)
JC-MB-LBP RGB	98.4 (-65.3)	90.5 (-41.9)	96.8 (-95.2)	100.0 (-100.0)	97.1 (-24.3)	90.8 (-68.4)	95.8 (-53.6)	✓ (−)
JC-MB-LBP LAB	100.0 (-57.3)	89.2 (-31.1)	96.8 (-96.0)	100.0 (-100.0)	98.2 (-12.0)	95.9 (-63.3)	97.1 (-45.8)	✓ (−)
JC-MB-LBP MR RGB	98.4 (-64.5)	91.9 (-43.2)	97.6 (-76.6)	100.0 (-100.0)	97.5 (-29.3)	92.9 (-75.5)	96.5 (-53.4)	✓ (−)
JC-MB-LBP MR LAB	100.0 (-47.6)	94.6 (-39.2)	96.0 (-93.5)	100.0 (-100.0)	98.6 (-20.7)	96.9 (-55.1)	97.8 (-46.8)	✓ (−)
LCVP $A^{(1)}$ RGB	93.5 (-61.3)	79.7 (-54.1)	91.1 (-87.9)	95.0 (-95.0)	84.1 (-22.5)	74.5 (-40.8)	85.5 (-48.3)	✓ (−)
LCVP $A^{(2)}$ RGB	91.9 (-72.6)	79.7 (-44.6)	84.7 (-66.1)	90.0 (-90.0)	84.8 (-15.6)	72.4 (-53.1)	83.9 (-44.4)	✓ (−)
LCVP $A^{(1)}$ LAB	96.0 (-64.5)	75.7 (-44.6)	87.1 (-85.5)	85.0 (-85.0)	83.3 (-25.4)	81.6 (-41.8)	85.2 (-48.5)	✓ (−)
LCVP $A^{(2)}$ LAB	88.7 (-58.9)	75.7 (-54.1)	88.7 (-82.3)	70.0 (-70.0)	81.9 (-12.7)	72.4 (-27.6)	82.0 (-40.6)	✓ (−)
LCVP MR $A^{(1)}$ RGB	96.0 (-62.9)	83.8 (-55.4)	90.3 (-89.5)	100.0 (-100.0)	89.9 (-19.2)	85.7 (-54.1)	90.1 (-49.7)	✓ (−)
LCVP MR $A^{(2)}$ RGB	94.4 (-66.1)	77.0 (-48.6)	88.7 (-71.8)	90.0 (-90.0)	88.0 (-15.2)	83.7 (-59.2)	87.6 (-45.4)	✓ (−)
LCVP MR $A^{(1)}$ LAB	96.0 (-65.3)	81.1 (-41.9)	91.1 (-90.3)	95.0 (-95.0)	89.5 (-23.2)	85.7 (-51.0)	89.7 (-49.9)	✓ (−)
LCVP MR $A^{(2)}$ LAB	91.1 (-58.9)	82.4 (-43.2)	91.1 (-75.0)	95.0 (-95.0)	84.8 (-20.3)	76.5 (-34.7)	85.9 (-42.9)	✓ (−)
GMRF k-NN	75.0 (-29.0)	63.5 (-56.8)	57.3 (-54.8)	30.0 (-30.0)	63.0 (-2.9)	69.4 (-54.1)	64.1 (-29.7)	✓ (−)
GMRF Bayes	91.9 (-50.8)	63.5 (-28.4)	68.5 (-59.7)	0.0 (0.0)	76.1 (-13.8)	82.7 (-60.2)	75.0 (-35.6)	✓ (−)
WT-GMRF k-NN	97.6 (-54.8)	83.8 (-71.6)	90.3 (-83.1)	80.0 (-70.0)	87.7 (-21.0)	87.8 (-60.2)	89.2 (-49.6)	✓ (−)
WT-GMRF Bayes	81.5 (-57.3)	23.0 (-17.6)	62.1 (-62.1)	0.0 (0.0)	94.6 (-7.2)	55.1 (-54.1)	71.2 (-32.7)	✓ (−)
WT-GMRF-CNH k-NN	95.2 (-54.8)	85.1 (-78.4)	90.3 (-82.3)	85.0 (-80.0)	88.8 (-16.7)	80.6 (-61.2)	88.5 (-48.9)	✓ (−)
WT-GMRF-CNH Bayes	78.2 (-66.1)	0.0 (0.0)	55.6 (-55.6)	0.0 (0.0)	98.9 (0.7)	31.6 (-31.6)	65.6 (-25.1)	✓ (−)

Table 4.6.: The detailed classification results obtained with the different methods evaluated on the pit pattern images (6 classes).

Celiac Disease Images

Figures 4.13 and 4.14 show the overall classification rates obtained with the different methods on the celiac disease image database. Again, the blue bars denote the results obtained with LOO-CV, while the the red bars denote the results obtained using LOPO-CV. The numbers on top of each pair of bars indicate the result drop when switching from LOO-CV to LOPO-CV. In addition we provide an indicator next to the method names, which – if filled gray – shows that an eventual result drop is statistically significant according to McNemar's test.

The detailed classification rates for the 2-classes case and the 4-classes case can be found in tables 4.7 and 4.8, respectively. In these tables we provide the detailed classification rates for the single classes in case of LOO-CV. In addition we provide the result drop when using LOPO-CV instead of LOO-CV for each result in brackets.

From figures 4.13 and 4.14 we notice that in case of LOO-CV the best overall classification rates in the 2-classes case has been achieved by the LCVP method when using multiple scales, followed by the WT-BBC method. When using four classes, again LCVP MR yields the highest overall classification rate. But the second-best method is JC-MB-LBP MR method this time. In case of LOPO-CV the picture is rather similar. Again the best overall classification rate in case of 2 classes is obtained by LCVP MR, followed by WT-BBC. In case of four classes there are two methods which perform equally well, LCVP MR and EDGEFEATURES, followed by WT-BBC.

As already observed in case of the pit pattern images, we also notice from figures 4.13 and 4.14 as well as from tables 4.7 and 4.8, that also in case of the celiac disease images the overall

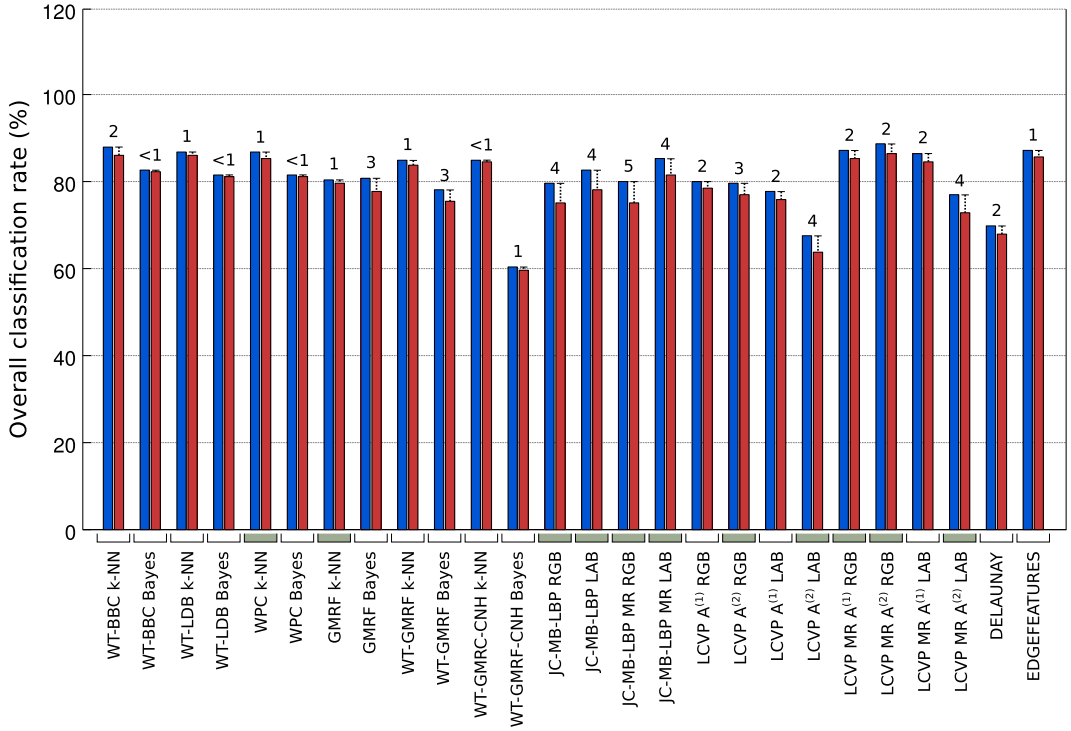


Figure 4.13.: Comparison of the overall classification rates for the celiac disease images where the blue and red bars denote the LOO-CV results and LOPO-CV results respectively (2 classes).

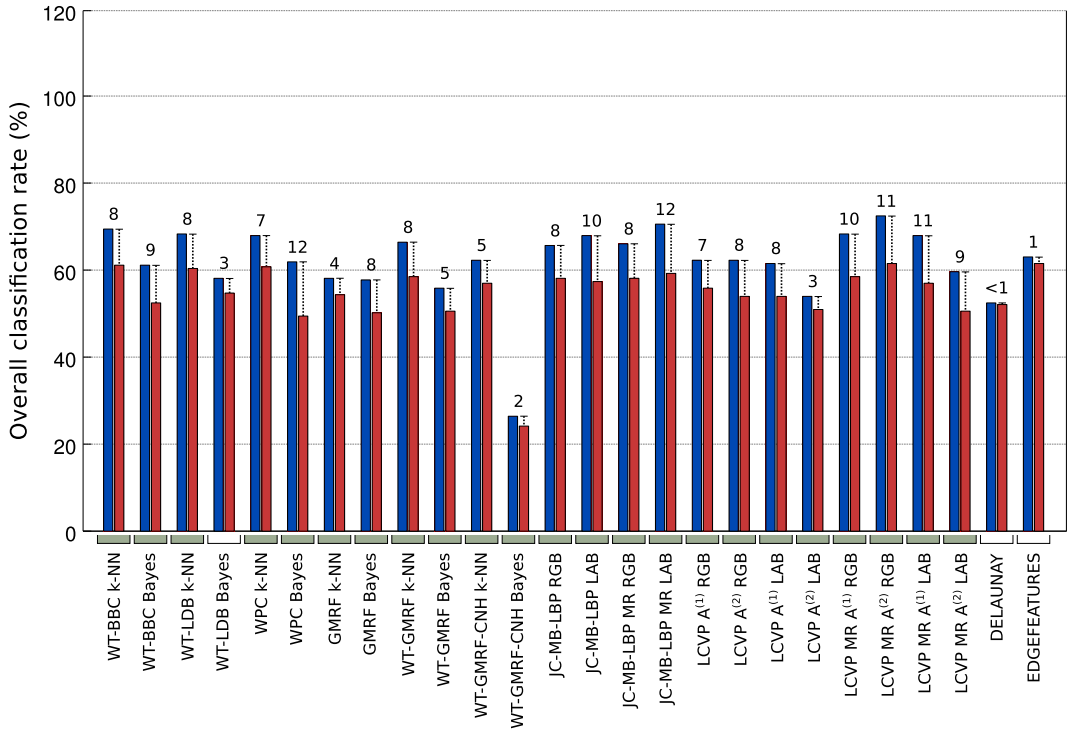


Figure 4.14.: Comparison of the overall classification rates for the celiac disease images where the blue and red bars denote the LOO-CV results and LOPO-CV results respectively (4 classes).

classification accuracies always drop, although the result drops are considerably lower as compared to the pit pattern image database. Moreover, in contrast to the pit pattern experiments, the result differences are sometimes statistically insignificant, which can be noticed clearly in the 2-classes case. As we also notice, the result drop when using LOPO-CV instead of LOO-CV is rather moderate, as compared to the pit pattern images, with values between less than 1% and 5% in the 2-classes case and between less than 1% and 12% in the 4-classes case. One reason for such moderate result drops is that the celiac disease image database uses about 3.6 images on average per patient, while in case of the pit pattern images we used about 17.9 images on average per patient. In addition, as already indicated in Section 4.2.2, in case of the pit pattern images LOPO-CV is very likely to unnecessarily reduce the training sets for different classes when classifying images from one specific class. This, however, is not possible in case of the celiac disease images, since in this database a patient is represented in a single class only.

Tables 4.7 and 4.8 reveal that in the 2-classes case none of both classes clearly exhibits the most result drops among the different methods when switching from LOO-CV to LOPO-CV. In the 4-classes case, however, class Marsh-0 seams to suffer least from LOPO-CV, while the remaining classes show result drops in roughly the same range. This may be attributed to the fact that in the 4-classes case class Marsh-0 contains significantly more images as compared to the classes Marsh-3A to Marsh-3C.

In figures 4.15 to 4.18 we again provide the method ranking for the different class counts as well as for the different cross-validation protocols used. As we notice, similar to the pit

Method	Marsh-0	Marsh-3	Total	S_{CV}
WT-LDB k-NN	88.2 (1.6)	85.6 (-3.6)	86.9 (-1.0)	
WT-LDB Bayes	73.9 (-0.7)	89.5 (0.0)	81.7 (-0.3)	
WT-BBC k-NN	89.2 (3.3)	86.9 (-7.2)	88.1 (-2.0)	
WT-BBC Bayes	78.1 (0.0)	87.3 (-0.7)	82.7 (-0.3)	
WPC k-NN	91.2 (-1.0)	82.4 (-2.0)	86.8 (-1.5)	✓ (−)
WPC Bayes	73.9 (-0.7)	89.5 (0.0)	81.7 (-0.3)	
DELAUNAY	75.2 (0.0)	64.7 (-4.2)	69.9 (-2.1)	
EDGEFEATURES	85.3 (-0.7)	88.9 (-2.3)	87.1 (-1.5)	
JC-MB-LBP RGB	90.2 (-0.3)	69.0 (-8.5)	79.6 (-4.4)	✓ (−)
JC-MB-LBP LAB	83.0 (-5.6)	82.4 (-3.3)	82.7 (-4.4)	✓ (−)
JC-MB-LBP MR RGB	94.8 (3.3)	65.7 (-13.7)	80.2 (-5.2)	✓ (−)
JC-MB-LBP MR LAB	89.5 (-1.0)	81.0 (-6.5)	85.3 (-3.8)	✓ (−)
LCVP $A^{(1)}$ RGB	82.7 (1.3)	77.8 (-4.9)	80.2 (-1.8)	
LCVP $A^{(2)}$ RGB	79.1 (-0.7)	80.4 (-4.6)	79.7 (-2.6)	✓ (−)
LCVP $A^{(1)}$ LAB	77.5 (-4.2)	78.4 (0.0)	77.9 (-2.1)	
LCVP $A^{(2)}$ LAB	58.5 (-1.6)	76.5 (-5.9)	67.5 (-3.8)	✓ (−)
LCVP MR $A^{(1)}$ RGB	83.7 (-3.3)	91.2 (-1.0)	87.4 (-2.1)	✓ (−)
LCVP MR $A^{(2)}$ RGB	83.3 (-2.0)	94.4 (-2.6)	88.9 (-2.3)	✓ (−)
LCVP MR $A^{(1)}$ LAB	82.0 (-3.6)	91.2 (-0.7)	86.6 (-2.1)	
LCVP MR $A^{(2)}$ LAB	66.3 (-6.9)	87.6 (-1.3)	77.0 (-4.1)	✓ (−)
GMRF k-NN	82.7 (-0.3)	78.4 (-1.3)	80.6 (-0.8)	✓ (−)
GMRF Bayes	77.1 (-7.2)	84.6 (0.7)	80.9 (-3.3)	
WT-GMRF k-NN	90.8 (-3.3)	78.8 (1.3)	84.8 (-1.0)	
WT-GMRF Bayes	69.0 (-2.6)	87.3 (-2.6)	78.1 (-2.6)	
WT-GMRF-CNH k-NN	86.6 (-1.3)	83.3 (0.7)	85.0 (-0.3)	
WT-GMRF-CNH Bayes	32.7 (0.0)	87.9 (-1.3)	60.3 (-0.7)	

Table 4.7.: The detailed classification results obtained with the different methods evaluated on the celiac disease images (2 classes).

pattern images, most methods are contained within the orange regions. This indicates that most methods perform significantly worse as compared to the best performing method. But, while in case of the pit pattern images the methods get shifted to the left when using LOPO-CV, for the celiac disease image database we notice that the methods get slightly shifted to the right. The fact that there is only a noticeable shift indicates that, as already pointed out above, the methods perform more stable as compared to the other methods when using LOPO-CV instead of LOO-CV.

While there is no method which clearly dominates the green regions, we notice that all methods contained within the green regions are based on the k-NN classifier. This suggests that in case of the celiac disease images the k-NN classifier mostly performs significantly better as compared to the Bayes classifier.

From these figures we also notice that the ordering of the methods is very similar, no matter whether we carry out a 2-class or 4-class classification. When comparing the plots for LOO-CV and LOPO-CV we notice that the ordering of the methods again is rather similar in both cases. This however is not surprising. Since, as we already saw above, switching from LOO-CV to LOPO-CV does not affect the results too much. Hence a similar ordering of the methods can be expected too. Nevertheless, at least in the 4-classes cases there are methods which seem to have profited from the overfitting in case of LOO-CV and, hence, perform considerably worse in case of LOPO-CV as compared to other methods (e.g. WPC Bayes in case of the 4-classes).

Method	Mars-0	Marsh-3A	Marsh-3B	Marsh-3C	Total	S_{CV}
WT-LDB k-NN	92.8 (-0.3)	42.1 (-5.3)	46.5 (-27.2)	43.3 (-13.4)	68.5 (-8.2)	✓ (−)
WT-LDB Bayes	83.7 (-9.5)	26.3 (33.7)	43.9 (-28.9)	24.7 (10.3)	58.0 (-3.3)	
WT-BBC k-NN	92.2 (-1.6)	38.9 (-15.8)	42.1 (-9.6)	58.8 (-19.6)	69.3 (-8.2)	✓ (−)
WT-BBC Bayes	82.4 (2.0)	34.7 (-17.9)	57.0 (-32.5)	23.7 (-4.1)	60.9 (-8.5)	✓ (−)
WPC k-NN	91.8 (0.7)	43.2 (-11.6)	42.1 (-11.4)	46.4 (-21.6)	67.8 (-7.0)	✓ (−)
WPC Bayes	74.5 (-4.6)	52.6 (-20.0)	65.8 (-22.8)	26.8 (-17.5)	61.9 (-12.4)	✓ (−)
DELAUNAY	92.8 (4.9)	7.4 (-6.3)	12.3 (-7.0)	16.5 (-4.1)	52.5 (-0.5)	
EDGEFEATURES	87.6 (3.3)	12.6 (-3.2)	62.3 (-14.0)	35.1 (1.0)	62.9 (-1.3)	
JC-MB-LBP RGB	90.2 (-0.3)	43.2 (-15.8)	49.1 (-14.0)	30.9 (-15.5)	65.8 (-7.7)	✓ (−)
JC-MB-LBP LAB	87.3 (-0.7)	61.1 (-20.0)	56.1 (-19.3)	26.8 (-20.6)	67.8 (-10.3)	✓ (−)
JC-MB-LBP MR RGB	96.4 (-4.9)	42.1 (-11.6)	43.0 (-14.0)	20.6 (-7.2)	66.0 (-8.0)	✓ (−)
JC-MB-LBP MR LAB	91.8 (0.0)	62.1 (-23.2)	57.0 (-23.7)	28.9 (-22.7)	70.8 (-11.6)	✓ (−)
LCVP $A^{(1)}$ RGB	87.6 (7.5)	32.6 (-11.6)	51.8 (-28.9)	24.7 (-19.6)	62.4 (-6.5)	✓ (−)
LCVP $A^{(2)}$ RGB	88.6 (1.0)	33.7 (-15.8)	41.2 (-17.5)	30.9 (-17.5)	62.1 (-8.0)	✓ (−)
LCVP $A^{(1)}$ LAB	89.2 (0.3)	30.5 (-8.4)	45.6 (-20.2)	23.7 (-17.5)	61.6 (-7.7)	✓ (−)
LCVP $A^{(2)}$ LAB	88.6 (0.0)	23.2 (-6.3)	28.9 (-7.9)	5.2 (-4.1)	54.1 (-3.1)	✓ (−)
LCVP MR $A^{(1)}$ RGB	91.2 (2.6)	48.4 (-27.4)	51.8 (-14.9)	34.0 (-24.7)	68.1 (-9.6)	✓ (−)
LCVP MR $A^{(2)}$ RGB	93.5 (-2.3)	49.5 (-22.1)	50.9 (-17.5)	53.6 (-18.6)	72.4 (-10.8)	✓ (−)
LCVP MR $A^{(1)}$ LAB	88.6 (2.3)	36.8 (-7.4)	64.0 (-36.8)	37.1 (-24.7)	67.8 (-10.8)	✓ (−)
LCVP MR $A^{(2)}$ LAB	80.1 (5.9)	33.7 (-20.0)	49.1 (-24.6)	34.0 (-28.9)	59.8 (-9.3)	✓ (−)
GMRF k-NN	88.9 (0.7)	22.1 (-4.2)	43.0 (-10.5)	14.4 (-9.3)	58.2 (-3.8)	✓ (−)
GMRF Bayes	96.7 (-1.6)	8.4 (-8.4)	36.0 (-24.6)	9.3 (-6.2)	57.8 (-7.7)	✓ (−)
WT-GMRF k-NN	91.8 (0.7)	43.2 (-14.7)	52.6 (-26.3)	25.8 (-6.2)	66.5 (-7.8)	✓ (−)
WT-GMRF Bayes	80.4 (19.3)	29.5 (-29.5)	43.9 (-43.0)	17.5 (-14.4)	55.7 (-5.2)	✓ (−)
WT-GMRF-CNH k-NN	91.8 (-1.6)	41.1 (-17.9)	35.1 (-7.9)	22.7 (-2.1)	62.4 (-5.4)	✓ (−)
WT-GMRF-CNH Bayes	17.3 (0.0)	71.6 (-1.1)	22.8 (-8.8)	14.4 (-2.1)	26.3 (-2.1)	✓ (−)

Table 4.8.: The detailed classification results obtained with the different methods evaluated on the celiac disease images (4 classes).

While in the 4-classes case there are also methods which seem to suffer from the overfitting and thus deliver a higher classification accuracy as compared to the other methods (e.g. the EDGEFEATURES method), most methods prove to be rather stable.

4.2.3. Ensemble Classification

In the previous sections we presented the results obtained with the different methods. In this section we present and discuss results obtained with our ensemble classifier as presented in [34].

An important issue in terms of the classifier ensemble is the combination of weak classifiers in order to obtain a combination performing better as compared to the weak methods. Since in our case we have 26 different method candidates at hand a manual selection would be a tedious task due to the high number of potential combinations ($2^{26} - 1 = 67.108.863$). Hence, we implemented a rather simple algorithm which, based on the overall classification rates of the single methods, determines a combination to be used for the classifier ensemble.

This algorithm starts by adding the method yielding the best overall classification rate to the ensemble method set. Then, based on a rating, new methods are added successively. After a new method has been added to the method set, the new set is rated according to a rating function. If the rating is below the rating without the newly added method, this method is

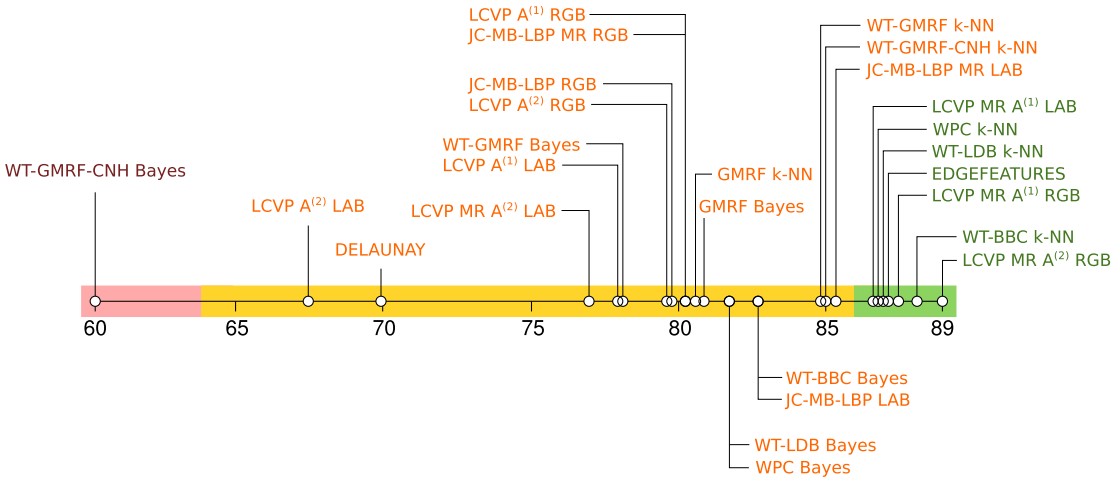


Figure 4.15.: Method ranking for the methods when applied to the celiac disease image database (2 classes, LOO-CV).

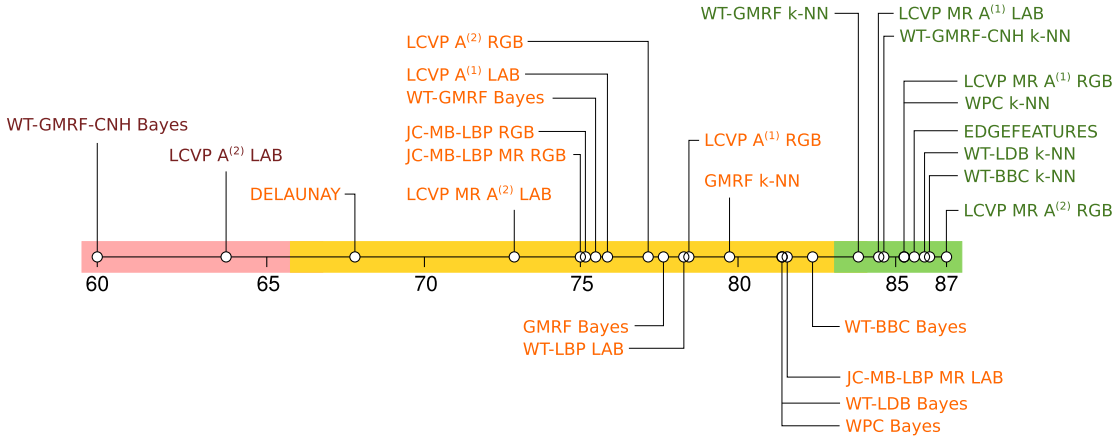


Figure 4.16.: Method ranking for the methods when applied to the celiac disease image database (2 classes, LOPO-CV).

removed from the set. The rating function compares the number of misclassifications for each image using the methods in the ensemble method set. If this number is below (above) the number of methods in the set divided by 2, the rating value is decremented (incremented). In other words, if more than half of the methods classify an image correctly, this is rewarded by incrementing the rating. Otherwise, a penalty is used to decrement the rating. If, however, the number of misclassifications is equal to the number of methods divided by 2 the rating remains untouched.

Since it is possible that, after the ensemble method set has been created by testing all methods available, adding an already tested method increases the rating, the algorithm carries out N iterations, where N corresponds to the total number of methods available for inclusion.

The method selection algorithm can be outlined as follows:

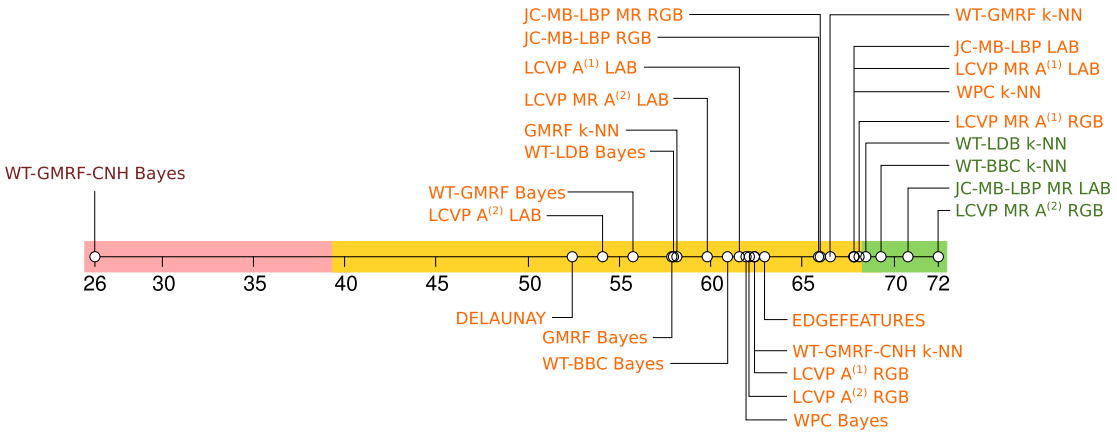


Figure 4.17.: Method ranking for the methods when applied to the celiac disease image database (4 classes, LOO-CV).

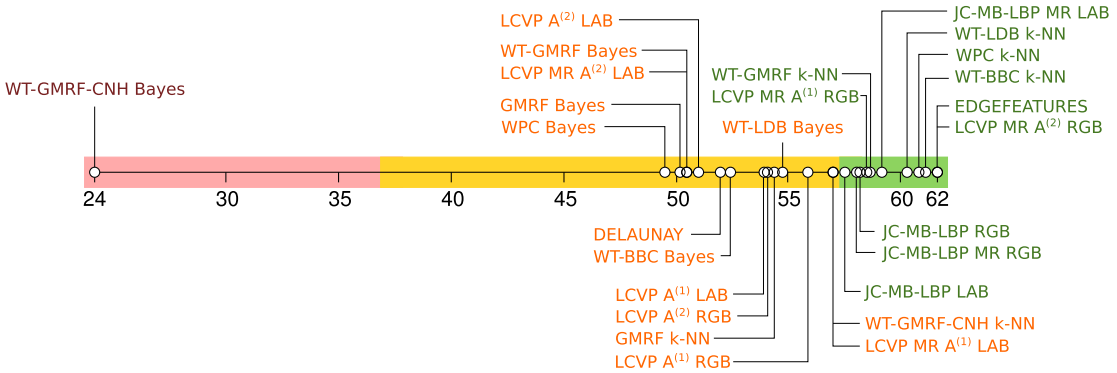


Figure 4.18.: Method ranking for the methods when applied to the celiac disease image database (4 classes, LOPO-CV).

1. Initialize the set of method candidates as $S = \{B\}$, where B denotes the method which yielded the highest overall classification rate.
2. Set ensemble set rating $R = 0$.
3. Set iteration counter to $I = 0$.
4. Initialize set of available methods as T with $T \cap S = \{\}$.
5. Set $S = S \cup M$ with $M \in T$ and $T = T \backslash M$ (in other words, M denotes the next available method which has not been added to S already).
6. Compute the rating value R_{new} by carrying out the following steps:
   a) Initialize new rating value with $R_{\text{new}} = 0$.
   b) For an image out of the image set compute the number of correct classifications among all methods in S as C.
   c) If $C < |S|/2$ then $R_{\text{new}} = R_{\text{new}} - 1$

d) If $C > |S|/2$ then $R_{\text{new}} = R_{\text{new}} + 1$

e) Continue with step 6b for each image in the image database.

7. If $R_{\text{new}} < R$ then $S = S \backslash M$.
8. If $R_{\text{new}} \geq R$ then $R = R_{\text{new}}$.
9. If $T \neq \{\}$ continue with step 5.
10. Increment iteration counter by setting $I = I + 1$.
11. If $I < N$ continue with step 4.

The results for the ensemble classification are shown in Table 4.9. The columns $\mathbf{R}_{\text{Best}}$, $\mathbf{R}_{\text{Ensemble}}$, and $\mathbf{S_E}$ denote the overall classification rate obtained with the best performing method within the method set S, the overall classification rate obtained with the ensemble classifier, and an indicator for statistical significance (according to McNemar's test with $\alpha = 0.05$) between the outcome of the best performing method in S and the ensemble (denoted by a ✓), respectively. In addition, we provide a "(+)" ("(-)") if the ensemble classifier achieved a higher (lower) overall classification accuracy as compared to the best performing method in S.

As we notice from this table in most cases the ensemble classifier yields at least slightly higher results as compared to the result achieved with the best performing method within the method set S. But we also see that the ensemble classifier performs significantly better only in one case (Pit pattern images, 2 classes, LOPO-CV). There also exist cases where the ensemble classifier performs equally well (Pit pattern images, 2 classes and 3 classes, LOO-CV) or even worse (Pit pattern images, 3 classes, LOPO-CV) as compared to the best single method in S.

The behavior that the ensemble classifier is in most cases not able to perform significantly better as compared to the methods in S can be explained by the fact that the selection algorithm outlined above always adds the best performing single method already in the first step. All other methods subsequently added usually yield an accuracy noticeably below the highest overall classification rate or a similar rate but with only insignificantly different method outcomes. As a consequence, either the poorly performing methods force down the ensemble classifier or the diversity between the methods in S is low.

Concerning the selected methods it is worthwhile to mention that the selection algorithm always picks out 3 methods for S. In all cases either the LCVP method or the JC-MB-LBP (both methods using multiple scales) has been chosen at least once. However, this is not surprising

	Pit Pattern Images					
	LOO-CV			**LOPO-CV**		
	$\mathbf{R}_{\text{Best}}$	$\mathbf{R}_{\text{Ensemble}}$	$\mathbf{S_E}$	$\mathbf{R}_{\text{Best}}$	$\mathbf{R}_{\text{Ensemble}}$	$\mathbf{S_E}$
2 classes	99.16	99.16		90.50	91.48	✓ (+)
3 classes	98.46	98.46		80.03	79.61	(−)
6 classes	97.77	98.18	(+)	51.26	52.23	(+)
	Celiac Disease Images					
	LOO-CV			**LOPO-CV**		
	$\mathbf{R}_{\text{Best}}$	$\mathbf{R}_{\text{Ensemble}}$	$\mathbf{S_E}$	$\mathbf{R}_{\text{Best}}$	$\mathbf{R}_{\text{Ensemble}}$	$\mathbf{S_E}$
2 classes	88.89	90.36	(+)	86.60	87.25	(+)
4 classes	72.39	75.00	(+)	61.60	62.09	(+)

Table 4.9.: Results of the ensemble classification compared to the best methods for each image database (LOO-CV and LOPO-CV).

since these methods are always performing best as we notice from tables 4.4 to 4.6 and tables 4.7 and 4.8.

In addition, in the case of the pit pattern images using LOPO-CV, the algorithm always selects methods based on different types of features (EDGEFEATURES method and the JC-MB-LBP method – either using a single scale or multiple scales). But also in case of the celiac disease images features from different domains are used (combinations of LCVP, JC-MB-LBP, EDGEFEATURES, and WT-BBC).

5. Conclusion and Future Outlook

As we have seen in the last chapters, throughout the last years we developed a set of different methods yielding different types of features (e.g. shape features or statistical features). We mainly evaluated these methods on endoscopic images in order to classify polyps, with the ultimate goal of detecting colon cancer. But we also evaluated these methods on endoscopic imagery with the aim of detecting celiac disease.

While the classification accuracies are very promising already, we noticed that at least in the case of the pit pattern images the classification accuracies heavily depend on the cross-validation method used. As a consequence, in case of LOO-CV the classification rates are comparable to the rates achieved by physicians [55]. However, in case of LOPO-CV the rates quite frequently significantly drop. For the celiac disease images the differences between these two cross-validation methods are not that pronounced. This must mainly be attributed to the fact that this image database contains far less subimages per original image, hence, the effect of overfitting is not that noticeable.

But there is also a potential to evaluate better suited features or optimize features developed in the past. In case of the DELAUNAY method, for example, we still suffer from ridges, as these are supposed to have a negative impact on the feature extraction process. Especially in case of this method it is necessary to make the pit candidate detection more robust in order to eliminate pit regions which are resulting from image noise.

In the past we also suffered from the imbalance in the pit pattern image database in terms of the images available per image class. This issue was also suspected to influence the recognition rates in a negative way. To alleviate this problem we have to emphasize a more well-balanced image set. This however could get problematic due to the difference prevalences of the different types of colonic lesions. In addition we have to evaluate ways to be able to deal better with imbalanced sets. It will also be important to collect enough images to be able to assemble different sets of images – one for the training and for the validation. This will also greatly improve the stability and explanatory power of the classification results achieved. In this course it would be interesting too – for pit pattern images as well as for celiac disease images – to use images captured using NBI in contrast to traditional chromo-endoscopy as NBI will replace the traditional technique more and more in future.

Another issue concerns the computational demand of the methods evaluated. While in our current scenario real-time capabilities were not the goal this might be of particular interest for future applications. Since for our work up to now we manually extracted regions of interest from the original endoscopic images, we nevertheless were restricted to an offline processing of the images. If, however, in future work we want to automatically extract regions of interest it will be imperative to lower the computation demand of the methods in order to be able to use them for a real-time scenario. Especially when considering that modern endoscopes provide HD resolution, this task gets even more challenging. There exist different potential options in order to lower the computational burden of our methods. Future research in this direction could therefore be to implement the methods with the help of GPUs on modern graphics cards. By utilizing the capabilities of GPUs in terms of massively parallel computing a considerable speed-up for our methods could be achieved (e.g. k-NN classifier for JC-MB-LBP or the edge extraction in case of our shape-based methods).

Bibliography

[1] A. A. Al-Rahayfeh and A. A. Abuzneid. Detection of bleeding in wireless capsule endoscopy images using range ratio color. *International Journal of Multimedia & Its Applications*, 2(2):1–10, May 2010.

[2] L. Alexandre, J. Casteleiro, and N. Nobre. Polyp detection in endoscopic video using SVMs. In *Knowledge Discovery in Databases: PKDD 2007*, pages 358–365. Springer Berlin, 2007.

[3] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilariño. Texture-based polyp detection in colonoscopy. In W. Brauer, H.-P. Meinzer, T. M. Deserno, H. Handels, and T. Tolxdorff, editors, *Bildverarbeitung für die Medizin 2009*, Informatik aktuell, pages 346–350. Springer Berlin, 2009.

[4] B. André, T. Vercauteren, A. Perchant, A. M. Buchner, M. B. Wallace, and N. Ayache. Endomicroscopic image retrieval and classification using invariant visual features. In *In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, pages 346–349, Boston, Massachusetts, USA, June 2009.

[5] B. André, T. Vercauteren, A. Perchant, M. B. Wallace, A. M. Buchner, and N. Ayache. Introducing space and time in local feature-based endomicroscopic image retrieval. In *Proceedings of the MICCAI 2009 Workshop – Medical Content-based Retrieval for Clinical Decision (MCBR-CDS'09)*, pages 18–30, London, UK, Sept. 2009.

[6] B. André, T. Vercauteren, M. B. Wallace, A. M. Buchner, and N. Ayache. Endomicroscopic video retrieval using mosaicing and visual words. In *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI'10)*, pages 1419–1422, Rotterdam, Netherlands, Apr. 2010.

[7] C. Barber, D. Dobkin, and H. Huhdanpaa. The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 2(4):469–483, Dec. 1996.

[8] 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 *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'09)*, pages 6683–6686, Minneapolis, Minnesota, USA, Sept. 2009.

[9] S. Bejakovic, R. Kumar, T. Dassopoulos, G. Mullin, and G. Hager. Analysis of crohn's disease lesions in capsule endoscopy images. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'09)*, pages 2793–2798, Kobe, Japan, May 2009.

[10] J. Bonnel, A. Khademi, S. Krishnan, and C. Ioana. Small bowel image classification using cross-co-occurrence matrices on wavelet domain. *Biomedical Signal Processing and Control*, 4(1):7–15, Jan. 2009.

[11] T. B. Borcharrt, A. Conci, and M. C. d'Ornellas. A warping based approach to correct distortions in endoscopic images. In *Proceedings of the 22nd Brazilian Symposium on Computer Graphics and Image Processing (Sibgrapi'09)*, Rio de Janeiro, Brazil, Oct. 2009.

[12] E. J. Candes and D. L. Donoho. Curvelets, multiresolution representation, and scaling laws. In *Proceedings of the SPIE – Wavelet Applications in Signal and Image Processing VIII*, volume 4119, pages 277–282, San Diego, CA, USA, July 2000.

[13] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, Sept. 1986.

[14] J. Church. Adenoma detection rate and the quality of colonoscopy: the sword has two edges. *Diseases of the Colon & Rectum*, 51(5):520–523, Mar. 2008.

[15] E. J. Ciaccio, C. A. Tennyson, S. K. Lewis, S. Krishnareddy, G. Bhagat, and P. H. Green. Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images. *Computer Methods and Programs in Biomedicine*, 100(1):39–48, Oct. 2010.

[16] R. R. Coifman and M. V. Wickerhauser. Entropy based methods for best basis selection. *IEEE Transactions on Information Theory*, 38(2):719–746, Mar. 1992.

[17] M. Coimbra, M. Mackiewicz, M. Fisher, C. Jamieson, J. Scares, and J. P. S. Cunha. Computer vision tools for capsule endoscopy exam analysis. *EURASIP Newsletter*, 18(1):1–19, Mar. 2007.

[18] K. Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4–5):192–211, 2007.

[19] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley & Sons, 2nd edition, 2000.

[20] B. Everitt. *The Analysis of Contingency Tables*. Chapman and Hall, 1977.

[21] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, Nov. 1984.

[22] H. Girgis, B. Mitchell, T. Dassopoulos, G. Mullin, and G. Hager. An intelligent system to detect chrohn's disease inflammation in wireless capsule endoscopy videos. In *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI'10)*, pages 1373–1376, Rotterdam, Netherlands, Apr. 2010.

[23] B. Giritharan, X. Yuan, J. Liu, B. Buckles, J. Oh, and S. J. Tang. Bleeding detection from capsule endoscopy videos. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'08)*, pages 4780–4783, Vancouver, British Columbia, Canada, Aug. 2008.

[24] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, 2007.

[25] S. Gross, M. Kennel, T. Stehle, J. Wul, J. Tischendorf, C. Trautwein, and T. Aach. Polyp segmentation in NBI colonoscopy. In W. Brauer, H.-P. Meinzer, T. M. Deserno, H. Handels, and T. Tolxdorff, editors, *Bildverarbeitung für die Medizin 2009*, Informatik aktuell, pages 252–256. Springer Berlin, 2009.

[26] S. Gross, T. Stehle, A. Behrens, R. Auer, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf. A comparison of blood vessel features and local binary patterns for colorectal polyp classification. In *Proceedings of Medical Imaging 2008: Computer-Aided Diagnosis*, volume 6918, Orlando, Florida, USA, Feb. 2009.

[27] M. Gschwandtner, M. Liedlgruber, A. Uhl, and A. Vécsei. Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, pages 1–6, Corfu, Greece, Nov. 2010.

[28] A. Häfner, A. Uhl, A. Vécsei, G. Wimmer, and F. Wrba. Complex wavelet transform variants and scale invariance in magnification-endoscopy image classification. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, Corfu, Greece, Nov. 2010.

[29] M. Häfner, L. Brunauer, H. Payer, R. Resch, A. Gangl, A. Uhl, A. Vécsei, and F. Wrba. Computer-aided classification of zoom-endoscopical images using fourier filters. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):958–970, July 2010.

[30] M. Häfner, L. Brunauer, H. Payer, R. Resch, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using DCT and FFT. In *Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, pages 159–164, Maribor, Slovenia, June 2007.

[31] M. Häfner, L. Brunauer, H. Payer, R. Resch, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopical colon images using evolved Fourier feature vectors. In *Proceedings of the 2007 IEEE Machine Learning for Signal Processing Workshop (MLSP'07)*, pages 99–104, Thessaloniki, Greece, Aug. 2007.

[32] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Combining Gaussian Markov random fields with the discrete wavelet transform for endoscopic image classification. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'09)*, pages 177–182, Santorini, Greece, July 2009.

[33] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using extended local binary patterns. In *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, pages 1–4, Larnaca, Cyprus, Nov. 2009.

[34] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using multichannel features and multiclassification. In D. F. T.P. Exarchos, A. Papadopoulos, editor, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, pages 335–350. IGI Global, Hershey, PA, USA, 2009.

[35] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Classification of endoscopic images using Delaunay triangulation-based edge features. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, volume 6112 of *Springer LNCS*, pages 131–140, Povoa de Varzim, Portugal, June 2010.

[36] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Endoscopic image classification using edge-based features. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, pages 2724–2727, Istanbul, Turkey, Aug. 2010.

[37] M. Häfner, C. Kastinger, H. Schmidt, K. Thonhauser, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Comparison of k-NN, SVM, and NN in pit pattern classification of zoom-endoscopic colon images using co-occurrence histograms. In *Proceedings of the 5th International Symposium on Image and Signal Processing and Analysis (ISPA'07)*, pages 516–521, Istanbul, Turkey, Sept. 2007.

[38] M. Häfner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using histogram techniques. In *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG'06)*, pages 58–61, Reykjavik, Iceland, June 2006.

[39] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei. Feature-extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. *Pattern Analysis & Applications*, 12(4):407–413, 2009.

[40] M. Häfner, R. Kwitt, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. One-against-one classification for zoom-endoscopy images. In *Proceedings of the 4th International Conference on Advances in Medical, Signal and Information Processing (MEDSIP'08)*, pages 1–4, Santa Margherita Ligure, Italy, July 2008.

[41] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Medical Image Analysis*, 16(1):75–86, 2012.

[42] M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using wavelet texture features. In W. Sandham, D. Hamilton, and C. James, editors, *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06)*, pages 1–4, Glasgow, Scotland, UK, July 2006.

[43] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621, 1973.

[44] S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vecsei. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 718–723, Salzburg, Austria, Sept. 2009.

[45] C.-R. Huang, P.-C. Chung, B.-S. Sheu, H.-J. Kuo, and M. Popper. Helicobacter pylori-related gastric histology classification using support-vector-machine-based feature selection. *IEEE Transactions Information Technology in Biomedicine*, 12(4):523–531, July 2008.

[46] D. K. Iakovidis, D. E. Maroulis, and S. A. Karkanis. An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy. *Computers in Biology and Medicine*, 36(10):1084–1103, Oct. 2006.

[47] D. K. Iakovidis, D. E. Maroulis, S. A. Karkanis, and A. Brokos. A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*, pages 575–580, Dublin, Ireland, June 2005.

[48] D. K. Iakovidis, D. E. Maroulis, S. A. Karkanis, P. Papageorgas, and M. Tzivras. Texture multichannel measurements for cancer precursors' identification using support vector machines. *Measurement*, 36(3-4):297–313, Oct. 2004.

[49] Y. S. Jung, Y. H. Kim, D. H. Lee, and J. H. Kim. Active blood detection in a high resolution capsule endoscopy using color spectrum transformation. In *Proceedings of the International Conference on BioMedical Engineering and Informatics (BMEI'08)*, pages 859–862, Sanya, Hainan, China, May 2008.

[50] J. Kang and R. Doraiswami. Real-time image processing system for endoscopic applications. In *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE'03)*, volume 3, pages 1469–1472, Niagara Falls, Ontario, Canada, May 2003.

[51] A. Karargyris and N. Bourbakis. A methodology for detecting blood-based abnormalities in wireless capsule endoscopy videos. In *Proceedings of the 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE'08)*, pages 1–6, Athens, Greece, Oct. 2008.

[52] A. Karargyris and N. Bourbakis. Identification of polyps in wireless capsule endoscopy videos using log gabor filters. In *Proceedings of the Life Science Systems and Applications Workshop (LiSSA'09)*, pages 143–147, Bethesda, MD, USA, Apr. 2009.

[53] S. A. Karkanis, D. Iakovidis, D. Karras, and D. Maroulis. Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures. In *Proceedings of the IEEE International Conference in Image Processing (ICIP'01)*, pages 833–836, Thessaloniki, Greece, Oct. 2001.

[54] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, Sept. 2003.

[55] S. Kato, K.-I. Fu, Y. Sano, T. Fujii, Y. Saito, T. Matsuda, I. Koba, S. Yoshida, and T. Fujimori. Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions. *World Journal of Gastroenterology*, 12(9):1416–1420, Mar. 2006.

[56] A. Khademi and S. Krishnan. Multiresolution analysis and classification of small bowel medical images. In *Proceedings of the 29th Annual Int Conf of the IEEE Engineering in Medicine and Biology Society (EMBS'07)*, volume 2007, pages 4524–4527, Vancouver, British Columbia, Canada, Aug. 2007.

[57] V. Kodogiannis. Computer-aided diagnosis in clinical endoscopy using neurofuzzy systems. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'04)*, pages 1425–1429, Budapest, Hungary, 2004.

[58] V. Kodogiannis, M. Boulougoura, J. Lygouras, and I. Petrounias. A neuro-fuzzy-based system for detecting abnormal patterns in wireless-capsule endoscopic images. *Neurocomputing*, 70:704–717, Jan. 2007.

[59] V. Kodogiannis and J. Lygouras. Neuro-fuzzy classification system for wireless-capsule endoscopic images. *International Journal of Electrical, Computer, and Systems Engineering*, 2:55–63, 2008.

[60] S. M. Krishnan, X. Yang, K. L.Chan, S. Kumar, and P. M. Y. Goh. Intestinal abnormality detection from endoscopic images. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'98)*, pages 895–898, Hong Kong, China, Oct. 1998.

[61] S. Kudo, S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori, and A. Yagyuu. Colorectal tumours and pit pattern. *Journal of Clinical Pathology*, 47:880–885, Oct. 1994.

[62] R. Kwitt and A. Uhl. Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of zoom-endoscopy images. In *Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'07)*, pages 1–8, Rio de Janeiro, Brasil, Oct. 2007.

[63] R. Kwitt and A. Uhl. Color eigen-subband features for endoscopy image classification. In *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'08)*, pages 589–592, Las Vegas, Nevada, United States, Apr. 2008.

[64] P. Y. Lau and P. Correia. Detection of bleeding patterns in WCE video using multiple features. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'07)*, pages 5601–5604, Vancouver, British Columbia, Canada, Aug. 2007.

[65] B. Li and M. Q. Meng. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Computers in Biology and Medicine*, 39(2):141–147, Feb. 2009.

[66] B. Li and M. Q. Meng. Texture analysis for ulcer detection in capsule endoscopy images. *Image and Vision Computing*, 27(9):1336–1342, Aug. 2009.

[67] B. Li and M. Q.-H. Meng. Analysis of the gastrointestinal status from wireless capsule endoscopy images using local color feature. In *Proceedings of the International Conference on Information Acquisition (ICIA'07)*, pages 553–557, Jeju City, Korea, July 2007.

[68] M. Liedlgruber and A. Uhl. Statistical and structural wavelet packet features for Pit pattern classification in zoom-endoscopic colon images. In P. Dondon, V. Mladenov, S. Impedovo, and S. Cepisca, editors, *Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07)*, pages 147–152, Arcachon, France, Oct. 2007.

[69] M. Liedlgruber and A. Uhl. Endoscopic image processing - an overview. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 707–712, Salzburg, Austria, Sept. 2009.

[70] M. Liedlgruber and A. Uhl. Predicting pathology in medical decision support systems in endoscopy of the gastrointestinal tract. In C. Jao, editor, *Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain*, pages 195–214. InTech, Rijeka, Croatia, 2011.

[71] M. Liedlgruber and A. Uhl. A summary of research targeted at computer-aided decision support in endoscopy of the gastrointestinal tract. Technical Report 2011-01, Department of Computer Sciences, University of Salzburg, Austria, http://www.cosy.sbg.ac.at/research/tr.html, 2011.

[72] M. Liedlgruber, A. Uhl, and A. Vécsei. Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'11)*, pages 1–6, Corfu, Greece, July 2011.

[73] C. S. Lima, D. Barbosa, J. R. A. Tavares, L. Monteiro, and L. Carvalho. Classification of endoscopic capsule images by using color wavelet features, higher order statistics and radial basis functions. In *Proceedings of the 30th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society (EMBS'08), pages 1242–1245, Vancouver, British Columbia, Canada, 2008.

[74] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[75] G. D. Magoulas, V. P. Plagianakos, and M. N. Vrahatis. Neural network-based colonoscopic diagnosis using on-line learning and differential evolution. *Applied Soft Computing*, 4(4):369–379, Apr. 2004.

[76] S. Mallat. *A Wavelet Tour Of Signal Processing*. Academic Press, 2nd edition, 1999.

[77] D. E. Maroulis, D. K. Iakovidis, S. A. Karkanis, and D. A. Karras. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer Methods and Programs in Biomedicine*, 70(2):151–66, Feb. 2003.

[78] M. Marsh. Gluten, major histocompatibility complex, and the small intestine. a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology*, 102(1):330–354, Jan. 1992.

[79] C. Münzenmayer, A. Kage, T. Wittenberg, and S. Mühldorfer. Computer-assisted diagnosis for precancerous lesions in the esophagus. *Methods of Information in Medicine*, 48:324–330, June 2009.

[80] G. Oberhuber, G. Granditsch, and H. Vogelsang. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *European Journal of Gastroenterology and Hepatology*, 11:1185–1194, Nov. 1999.

[81] T. Ojala and M. Pietikäinen. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3):477–486, Mar. 1999.

[82] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, Jan. 1996.

[83] G. Pan, G. Yan, X. Qiu, and J. Cui. Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. *Journal of Medical Systems*, pages 1–8, Jan. 2010.

[84] B. Penna, T. Tillo, M. Grangetto, E. Magli, and G. Olmo. A technique for blood detection in wireless capsule endoscopy images. In *Proceedings of the 17th European Signal Processing Conference (EUSIPCO'09)*, pages 1864–1868, Glasgow, Scotland, 2009.

[85] J. Pesquet, H. Krim, and H. Carfantan. Time invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44(8):1964–1970, Aug. 1996.

[86] N. Saito and R. R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5(4):337–358, Dec. 1995.

[87] S. Seguí, L. Igual, F. Vilariño, P. Radeva, C. Malagelada, F. Azpiroz, and J. Vitrià. Diagnostic system for intestinal motility disfunctions using video capsule endoscopy. In A. Gasteratos, M. Vincze, and J. Tsotsos, editors, *Computer Vision Systems*, volume 5008 of *LNCS*, pages 251–260. Springer Berlin, 2008.

[88] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury. The dual-tree complex wavelet transform - a coherent framework for multiscale signal and image processing. *IEEE Signal Processing Magazine*, 22(6):123–151, Nov. 2005.

[89] S. Smith and J. Brady. SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[90] A. Sousa, M. Dinis-Ribeiro, M. Areia, and M. Coimbra. Identifying cancer regions in vital-stained magnification endoscopy images using adapted color histograms. In *Proceedings of the 16th International Conference on Image Processing (ICIP'09)*, pages 681–684, Cairo, Egypt, Nov. 2009.

[91] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf. Classification of colon polyps in NBI endoscopy using vascularization features. In *Medical Imaging 2009: Computer-Aided Diagnosis*, volume 7260, Orlando, Florida, USA, Feb. 2009. SPIE.

[92] P. M. Szczypinski, R. D. Sriram, P. V. J. Sriram, and D. N. Reddy. A model of deformable rings for interpretation of wireless capsule endoscopic videos. *Medical Image Analysis*, 13(2):312–324, Apr. 2009.

[93] T. Tamaki, J. Yoshimuta, T. Takeda, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, and S. Tanaka. A system for colorectal tumor classification in magnifying endoscopic nbi images. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision – ACCV 2010*, volume 6493 of *LNCS*, pages 452–463. Springer Berlin, 2011.

[94] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and Modelling of Faces and Gestures*, volume 4778 of *LNCS*, pages 168–182. Springer, Oct. 2007.

[95] J. J. W. Tischendorf, S. Gross, R. Winograd, H. Hecker, R. Auer, A. Behrens, C. Trautwein, T. Aach, and T. Stehle. Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study. *Endoscopy*, 42(3):203–207, Mar. 2010.

[96] M. P. Tjoa and S. M. Krishman. Feature extraction for the analysis of colon status from the endoscopic images. *Biomedical Engineering Online*, Apr. 2003.

[97] A. Vécsei, G. Amann, S. Hegenbart, M. Liedlgruber, and A. Uhl. Automated marsh-like classification of celiac disease in children using an optimized local texture operator. *Computers in Biology and Medicine*, 41(6):313–325, June 2011.

[98] A. Vécsei, T. Fuhrmann, M. Liedlgruber, L. Brunauer, H. Payer, and A. Uhl. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. *Computer Methods and Programs in Biomedicine*, 95:S68–S78, 2009.

[99] A. Vécsei, T. Fuhrmann, and A. Uhl. Towards automated diagnosis of celiac disease by computer-assisted classification of duodenal imagery. In *Proceedings of the 4th International Conference on Advances in Medical, Signal and Information Processing (MEDSIP'08)*, pages 1–4, Santa Margherita Ligure, Italy, 2008.

[100] E. Wadge, M. Boulougoura, and V. Kodogiannis. Computer-assisted diagnosis of wireless-capsule endoscopic images using neural network based techniques. In *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA'05)*, pages 328–333, Taormina, Sicily, Italy, July 2005.

[101] L. Wang and D. He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910, Aug. 1990.

[102] K. Zuiderveld. Contrast limited adaptive histogram equalization. In P. S. Heckbert, editor, *Graphics Gems IV*, pages 474–485. Morgan Kaufmann, 1994.

A. Appendix

A.1. Breakdown of Authors' Contribution

In the following we break the contribution of the authors, who contributed to the different publications, down. In case of equal contribution, the author names appear in alphabetical order on the publications.

Andreas Uhl is or was thesis advisor/project leader of Leonhard Brunauer, Thomas Fuhrmann, Michael Gschwandtner, Sebastian Hegenbart, Roland Kwitt, Michael Liedlgruber, and Hannes Payer. Since the explicit contribution of an advisor and project leader cannot be stated for a single paper, it is omitted in the following breakdown.

The medical experts who contributed image material and/or histopathological results to this thesis are Gabrielle Amann, Alfred Gangl, Michael Häfner, Andreas Vécsei, and Friedrich Wrba. The contribution of these authors is restricted to the medical background and material used throughout our publications. Since the breakdown concerns with the technical part of our publications only, the contribution of the medical experts is omitted in the following breakdown.

Publication	Contribution (in %)						
	Leonhard Brunauer	Thomas Fuhrmann	Michael Gschwandtner	Sebastian Hegenbart	Roland Kwitt	Michael Liedlgruber	Hannes Payer
M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl. Pit pattern classification of zoom-endoscopic colon images using wavelet texture features. In W. Sandham, D. Hamilton, and C. James, editors, *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06)*, pages 1–4, Glasgow, Scotland, UK, July 2006						100	
M. Liedlgruber and A. Uhl. Statistical and structural wavelet packet features for Pit pattern classification in zoom-endoscopic colon images. In P. Dondon, V. Mladenov, S. Impedovo, and S. Cepisca, editors, *Proceedings of the 7th WSEAS International Conference on Wavelet Analysis & Multirate Systems (WAMUS'07)*, pages 147–152, Arcachon, France, Oct. 2007						100	

Publication	Contribution (in %)						
	Leonhard Brunauer	Thomas Fuhrmann	Michael Gschwandtner	Sebastian Hegenbart	Roland Kwitt	Michael Liedlgruber	Hannes Payer
M. Liedlgruber and A. Uhl. Endoscopic image processing - an overview. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 707–712, Salzburg, Austria, Sept. 2009						100	
S. Hegenbart, R. Kwitt, M. Liedlgruber, A. Uhl, and A. Vecsei. Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA'09)*, pages 718–723, Salzburg, Austria, Sept. 2009				60	20	20	
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using multichannel features and multiclassification. In D. F. T.P. Exarchos, A. Papadopoulos, editor, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, pages 335–350. IGI Global, Hershey, PA, USA, 2009						100	
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Combining Gaussian Markov random fields with the discrete wavelet transform for endoscopic image classification. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'09)*, pages 177–182, Santorini, Greece, July 2009						100	
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Pit pattern classification using extended local binary patterns. In *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, pages 1–4, Larnaca, Cyprus, Nov. 2009						100	

Publication	Contribution (in %)						
	Leonhard Brunauer	Thomas Fuhrmann	Michael Gschwandtner	Sebastian Hegenbart	Roland Kwitt	Michael Liedlgruber	Hannes Payer
A. Vécsei, T. Fuhrmann, M. Liedlgruber, L. Brunauer, H. Payer, and A. Uhl. Automated classification of duodenal imagery in celiac disease using evolved fourier feature vectors. *Computer Methods and Programs in Biomedicine*, 95:S68–S78, 2009	25	25				25	25
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Classification of endoscopic images using Delaunay triangulation-based edge features. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, volume 6112 of *Springer LNCS*, pages 131–140, Povoa de Varzim, Portugal, June 2010						100	
M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Endoscopic image classification using edge-based features. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, pages 2724–2727, Istanbul, Turkey, Aug. 2010						100	
M. Gschwandtner, M. Liedlgruber, A. Uhl, and A. Vécsei. Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB'10)*, pages 1–6, Corfu, Greece, Nov. 2010			20			80	

Publication	Contribution (in %)						
	Leonhard Brunauer	Thomas Fuhrmann	Michael Gschwandtner	Sebastian Hegenbart	Roland Kwitt	Michael Liedlgruber	Hannes Payer
M. Liedlgruber and A. Uhl. Predicting pathology in medical decision support systems in endoscopy of the gastrointestinal tract. In C. Jao, editor, *Efficient Decision Support Systems – Practice and Challenges in Biomedical Related Domain*, pages 195–214. InTech, Rijeka, Croatia, 2011						100	
M. Liedlgruber, A. Uhl, and A. Vécsei. Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In *Proceedings of the 17th International Conference on Digital Signal Processing (DSP'11)*, pages 1–6, Corfu, Greece, July 2011						100	
A. Vécsei, G. Amann, S. Hegenbart, M. Liedlgruber, and A. Uhl. Automated marsh-like classification of celiac disease in children using an optimized local texture operator. *Computers in Biology and Medicine*, 41(6):313–325, June 2011				80		20	
M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Medical Image Analysis*, 16(1):75–86, 2012						100	
M. Liedlgruber and A. Uhl. A summary of research targeted at computer-aided decision support in endoscopy of the gastrointestinal tract. Technical Report 2011-01, Department of Computer Sciences, University of Salzburg, Austria, `http://www.cosy.sbg.ac.at/research/tr.html`, 2011						100	