Predicting the Histology of Colorectal Lesions in a Probabilistic Framework

Roland Kwitt and Andreas Uhl*
Department of Computer Sciences, University of Salzburg, Austria
{rkwitt,uhl}@cosy.sbg.ac.at

Michael Häfner
Department for Internal Medicine, St. Elisabeth Hospital, Austria

Alfred Gangl and Friedrich Wrba
Department of Gastroenterology/Clinical Pathology, Medical University of Vienna, Austria

Andreas Vécsei
St. Anna Children's Hospital, Austria

Abstract

In this paper, we present a novel approach to predict the histological diagnosis of colorectal lesions from high-magnification colonoscopy images by means of Pit Pattern analysis. Motivated by the shortcomings of discriminant classifier approaches, we present a generative model based strategy which is closely related to content-based image retrieval (CBIR) systems. The ingredients of the approach are the Dual-Tree Complex Wavelet Transform (DTCWT) and the mathematical construct of copulas. Our experimental study on a set of 627 *images confirms, that the joint statistical model leads to impressive prediction results compared to previous work.*

1. Motivation

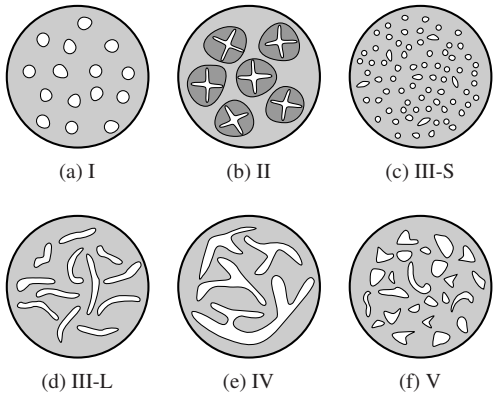
According to the statistics of the American Cancer Society [1], colorectal cancer is the third most commonly diagnosed cancer and the third leading cause of US cancer deaths in both men and women. Colorectal cancer is a paramount example where existing knowledge in combination with early screening procedures can prevent death and save lives. Computer-aided diagnosis systems have gained a lot of research interest recently. A lot of work has been done on the automated discrimination between normal and cancerous tissue using microscopic imaging, mainly by means of texture analysis [5, 31]. While these studies work directly with tissue samples of resected specimen obtained from biopsies, other works have studied the versatility of endoscopic video-frame processing for the detection of colorectal polyps [14, 26] and the assessment of colorectal abnormalities [20, 15, 12, 11]. However, conventional white-light video colonoscopy as it is used in these studies has its limitations, especially with respect to the detection of flat and depressed lesions [10]. The emergence of high-magnification chromoscopic colonoscopy (HMCC) poses several advantages over white-light video colonoscopy. In HMCC, high-magnification endoscopes with zoom-factors of up to $150\times$ are used to visualize the appearance of the colon mucosa. The high optical zoom and resolution reveal characteristic surface patterns (*i.e*. Pit Patterns) which can be analyzed by the experienced physician to predict the histological diagnosis. This visual inspection is guided by the Kudo criteria for Pit Pattern analysis. Usually, chromoagents such as indigo-carmine or methylene-blue are used during endoscopic examination to enhance the visual appearance of the observed tissue. As a matter of fact, HMCC is suggested as an *in vivo* staging tool to enhance the diagnostic process and guide therapeutic strategies.

1.1. Pit Pattern Analysis

Colorectal cancer predominantly develops from adenomatous polyps (adenomas), although adenomas do not inevitably become cancerous. Polyps of the colon are a frequent finding and are usually divided into metaplastic, adenomatous or malignant. Since the resection of all polyps is rather time-consuming, it is imperative that those polyps which warrant resection can be distinguished. The classification scheme presented by Kudo *et al*. [21] divides the mucosal crypt patterns into five types (Pit Patterns I–V). Fig. 1 provides a schematic illustration of the different Pit

*This work is funded by the Austrian Science Fund (FWF) under Project No. L366-N15.

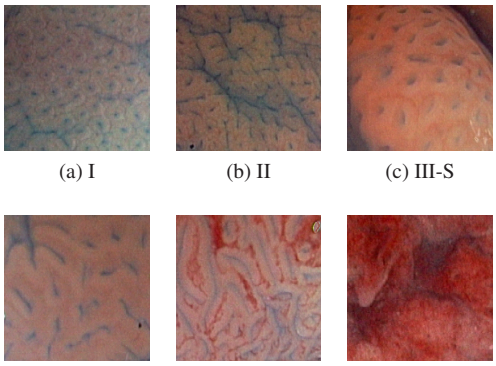
[1] http://www.cancer.org (accessed March 19th, 2010)



(a) I (b) II (c) III-S

(d) III-L (e) IV (f) V

Figure 1: Schematic illustration of the six colorectal crypt architectures, according to the Kudo criteria [21].



(a) I (b) II (c) III-S

(d) III-L (e) IV (f) V

Figure 2: Representative HMCC images of the different Pit Patterns. Note that types I and II show non-neoplastic lesions while types III-S, III-L, IV and V show neoplastic disease.

Patterns and Table 1 gives a textual description of their visual appearance. Exemplary HMCC images are shown in Fig. 2. While Pit Patterns I and II are characteristic of benign lesions and represent normal colon mucosa or hyperplastic polyps (*i.e.* non-neoplastic lesions), Pit Patterns III to V represent adenomatous and carcinomatous structures (*i.e.* neoplastic lesions). This includes serrated, tubular and tibulovillous adenoma as well as lymphoma, carcinoma and adenocarninoma.

At first sight, the Kudo criteria seems to be straightforward and easy to be applied. Nevertheless, it needs some experience and exercising to achieve good results. Correct diagnosis very much relies on the experience of the gastroenterologist as the interpretation of the Pit Patterns may be challenging [9]. Computer-assisted diagnosis is motivated by the work of Kato *et al.* [17], where the authors state that assessing the type of mucosal crypt patterns can actually predict the histological findings to a very high accuracy.

Pit Pattern	Visual Appearance
I	Round pit (normal pit)
II	Asteroid pit, stellar or papillary
III-S	Tubular or round pit (smaller than I)
III-L	Tubular or round pit (larger than I)
IV	Dentritic or gyrus-like pit
V	Irregular arrangement and sizes

Table 1: Description of the visual appearance of the colorectal crypt patterns observed during HMCC.

Regarding the correlation between the Pit Patterns and the histological findings, several (human-based) studies report good results for distinguishing non-neoplastic from neoplastic lesions, although with different diagnostic accuracies. A recent comparative study by Kato *et al.* [16] reports a histological prediction accuracy of 99.1% by means of HMCC and Pit Pattern analysis. Hurlstone *et al.* [10] claim a rate of approximately 95%, Tung *et al.* [33] claim 80.1%, however at a very low sensitivity of only 64.6%. In another work, Fu *et al.* [6] report 95.6% for HMCC compared to 84.0% using conventional white-light colonoscopy and 89.3% using chromoendoscopy without magnification. An even larger spread between HMCC and conventional white-light colonoscopy is listed by Konishi *et al.* [18] with 92% and 68%, respectively. In addition, inter-observer variability of HMCC-based diagnosis has been described at least for Barret's esophagus [27]. This inter-observer variability may to a lesser degree be also present in the interpretation of Pit Patterns of colonic lesions.

1.2. Objective

The objective of this work is two-fold: first, we want to reliably discriminate Pit Patterns I and II from III to V which amounts to identify non-neoplastic and neoplastic lesions. According to the medical literature, this is the clinically most relevant application scenario of the Pit Pattern analysis scheme. We denote this problem as the *two-class* problem. Second, we focus on a more therapeutically relevant subcategorization (see [16, 10]) in which neoplastic lesions are further discriminated into invasive and non-invasive types. The subgroups differ in that the representatives of the non-invasive class allow endoscopic mucosal resection (EMR), whereas representatives of the invasive class may require surgical resection. This second problem will be denoted as *three-class* problem. We adhere to the particular class assignment of Hurlstone *et al.* [10] where Pit Patterns III-L and IV are grouped into the non-invasive class and Pit Patterns III-S and V represent the invasive class.

2. A Generative Model Based Approach

In the context of our prediction/classification problem, we identify three critical issues related to discriminant classifier approaches: first, classifier training usually requires a sufficiently large number of training samples. Unless this can be guaranteed, we will inevitably run into overtraining issues. Second, most classifiers additionally require balanced class distributions. Unfortunately, we cannot guarantee this requirement either. Since some Pit Patterns (*e.g.* III-S) occur very rarely, the image distribution tends to be highly unbalanced. Neglecting this fact leads to overtraining in favor of classes with a large number of samples (see [28]). Third, we want to ensure that images with an already assigned histopathological diagnosis can be added to the image database at any time without effort. This avoids presumably time-consuming and unnecessary maintenance operations which might prevent the actual deployment in clinical practice. Since discriminant classifiers usually need re-training in case new samples are added, this requirement cannot be met as well.

2.1. The Probabilistic Framework

As a possible solution to the aforementioned disadvantages, we propose to employ a prediction strategy based on generative models. The baseline for this proposal is the framework of Bayesian image retrieval [34], also referred to as *Minimum Probability of Error* retrieval. Considering the medical image classification/prediction problem from the viewpoint of image retrieval brings along several advantages which correspond to the requirements stated above. An unknown HMCC image is considered as a query image in the probabilistic framework and classification is performed by first searching for the most similar image in the database of available HMCC images with an assigned histological diagnosis. Then, the class of the retrieved image is used as a prediction for the class of the unknown image. In classification terminology, this resembles a nearest neighbor classifier.

Next, we briefly recapitulate the theoretical foundation of the probabilistic formulation of CBIR which serves as the basis for our work. We assume that we have L database images $\mathcal{I}_1, \ldots, \mathcal{I}_L$. In the formulation of Vasconcelos & Lippman, an image $\mathcal{I}$ consists of a number of pixel observations $(x_1, \ldots, x_N) = \boldsymbol{x} \in \mathcal{X}$ residing in the space of observations $\mathcal{X}$. Further, each image belongs to its own image class and the classes have equal prior probability. The random variable Y signifies the class membership, with probability mass function $p_Y(y) = 1/L$. In a first step, referred to as the *feature transformation* step, we map an image from the space of observations to the so called *feature space* $\mathcal{Z}$, *i.e.* $T : \mathcal{X} \rightarrow \mathcal{Z}$. Consequently, $\boldsymbol{z} = T(\boldsymbol{x})$ denotes a so called *feature vector*. The key issue here is, to represent the image content in a domain which is more suitable for further processing. In the second step, a probabilistic model is constructed which captures how the feature vectors populate the feature space with respect to their class membership. The corresponding class-conditional p.d.f. $p_{\boldsymbol{Z}|Y}(\boldsymbol{z}|y)$ constitutes the *feature representation*. In the final step, we identify a *retrieval function* which retrieves the most similar image from the database and assigns the corresponding class label to the unknown image. This retrieval function $g : \mathcal{Z} \rightarrow \{1, \ldots, L\}$ is designed to minimize the probability of retrieval error. In [34], it is shown that the optimum retrieval function for this criterion is given by

$$g(\boldsymbol{z}) = \arg\max_y p_{\boldsymbol{Z}|Y}(\boldsymbol{z}|y) p_Y(y). \tag{1}$$

By noting that $p_Y(y)$ is a constant term, the retrieval function is simply the Maximum-Likelihood (ML) selection criterion. In any practical scenario, we have to estimate $p_{\boldsymbol{Z}|Y}$ from a collection of feature vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M$ and the actual retrieval process will be based on a collection of feature vectors $\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_K^*$, extracted from the unknown image $\mathcal{I}^*$. The assumption that the feature vectors are i.i.d. and conditionally independent given the true class membership facilitates estimation of $p_{\boldsymbol{Z}|Y}$ and allows to write the ML selection rule as

$$g(\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_K^*) = \arg\max_y \prod_{k=1}^{K} p_{\boldsymbol{Z}|Y}(\boldsymbol{z}_k^*|y). \tag{2}$$

We omit the notation $Z|Y$ from this point on and instead indicate that a feature representation belongs to image $\mathcal{I}_j$ by adding the superscript j to the collection of model parameters $\boldsymbol{\Theta}^{(j)}$.

Regarding the choice of feature transformation, we build upon previous research work [22] and use the Dual-Tree Complex Wavelet Transform (DTCWT). This transformation has turned out to be beneficial for image analysis purposes, since it provides high directional selectivity and approximate shift-invariance at low computational cost, see [30]. In contrast to the pyramidal DWT, we get six complex orientation subbands per decomposition scale. In the following section, we discuss the contribution of this work, namely the feature representation for the complex transform coefficient magnitudes of all subbands (from all color channels) on *one* decomposition level.

2.2. A Copula-based Feature Representation

Many previous approaches in the field of image retrieval, *e.g.* [3, 24], solely focus on statistical models for the marginal distributions of wavelet coefficients, mainly due to the associated computational advantages. However, we expect that additional information about the association structure of wavelet coefficients across subbands and color channels plays a key part in contributing to higher discrimination

rates. Unfortunately, many multivariate statistical models tend to become analytically quite involved, especially when it comes to parameter estimation and it is in general not possible to have marginal distributions from different distribution families. An elegant way to cover both problems is to rely on the mathematical construct of copulas. From a formal point of view, a copula is a B-dimensional distribution function $C : [0,1]^B \rightarrow [0,1]$ with uniform marginals, satisfying certain regularity conditions, see [29]. The key element of copula theory is Sklar's theorem [32]. Given a B-dimensional distribution function $F_{\mathbf{X}}$ of a random vector $\mathbf{X}$ with marginals $F_1, \ldots, F_B$, it states that there exists a B-dimensional copula C such that

$$F_{\mathbf{X}}(x_1, \ldots, x_B) = C(F_1(x_1), \ldots, F_B(x_B)), \quad (3)$$

exploiting the fact that every random variable can be transformed to a uniform random variable by its probability integral transform, *i.e.* the mapping $\mathbb{R}^B \rightarrow [0,1]^B, (x_1, \ldots, x_B) \mapsto (F_1(x_1), \ldots, F_B(x_B))$. Since we only consider the case of random vectors $\mathbf{X} = (X_1, \ldots, X_B)$ with continuous and strictly increasing marginal distribution functions, the copula is uniquely determined on $[0,1]^B$. As a corollary of Sklar's theorem, it follows that given a B-dimensional distribution function F with margins $F_1, \ldots, F_B$ and copula C, we have the relation

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), \ldots, F_B^{-1}(u_B)) \quad (4)$$

where F_i^{-1} denotes the quantile function of the i-th margin and $\mathbf{u} \sim U([0,1]^B)$. Regarding the process of finding a suitable statistical model for a collection of multivariate observations, the copula framework brings along an appealing simplification: the process of modeling the marginals is completely decoupled from the process of modeling the association structure. This allows to thoroughly adopt the findings of previous research work on the marginal distributions of complex wavelet coefficient magnitudes, however, raises the question which copula to choose.

In [25], it was shown that the Student t and the Gaussian copula are suitable choices to capture the association structure among the coefficients. Both copulas are members of the family of elliptical copulas and arise from the family of elliptical distributions. In fact, they *are* the copulas of elliptical distributions and inherit all the properties such as simple simulation of random numbers or well-known parameter estimation procedures for example. Since estimation of the Student t copula is computationally more involved than estimation of the Gaussian copula, we focus on the latter choice here. Further, we will later see that it is computationally less expensive to evaluate the likelihood of a collection of multivariate observations under a Gaussian copula model.

As we have noted before, the DTCWT leads to six complex detail subbands per decomposition level. Considering all subbands on a specific scale of a decomposed color image gives $B = 18$ subbands in total. Consequently, a feature vector $\mathbf{z}$ contains $B = 18$ elements, where each element is a transform coefficient magnitude $z_i = |z_i|$ from one subband, *i.e.* $\mathbf{z} = (z_1, \ldots, z_B)$. The joint p.d.f. of the copula-based model can be written as

$$p_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\Theta}) = c(F_1(z_1; \boldsymbol{\theta}_1), \ldots, F_B(z_B; \boldsymbol{\theta}_B); \boldsymbol{\Lambda}) \cdot \prod_{i=1}^{B} f_i(z_i; \boldsymbol{\theta}_i) \quad (5)$$

where c denotes the copula p.d.f., f_i denotes the p.d.f. of the i-th margin and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_B, \boldsymbol{\Lambda}\}$. In our setup, the copula is restricted to a Gaussian copula and the marginal distributions F_i are limited to two-parameter Weibull distributions with shape $\alpha_i > 0$ and scale $\beta_i > 0$, *i.e.* $\boldsymbol{\theta}_i = [\alpha_i \; \beta_i]$. The p.d.f. of a Gaussian copula is determined in the next section.

2.2.1 Parameter Estimation

Due to the fact, that it is computationally expensive and numerically cumbersome to jointly estimate the parameters of the marginal distributions and the copula parameters (denoted as the exact Maximum-Likelihood approach), we follow a convenient two-step procedure, termed the *Inference Functions from Margins (IFM)* method. The basic idea was introduced by Joe [13] and is based on a very simple decoupling of the estimation procedure. Given a collection of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M$ observations, we first estimate the parameters of the margins (*e.g.* Weibull)

$$\forall n \in \{1, \ldots, B\} : \hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{M} \log f_n(z_{in}; \boldsymbol{\theta}) \quad (6)$$

using ML estimation, see [19]. Second, we use the obtained estimates to conduct a probability integral transform on the margins, *i.e.* $y_{ij} = F_j(z_{ij}; \boldsymbol{\theta}_j)$. Third, we estimate the copula parameters in a ML sense by

$$\hat{\boldsymbol{\Lambda}} = \arg\max_{\boldsymbol{\Lambda}} \sum_{i=1}^{M} \log c(y_{i1}, \ldots, y_{iB}; \boldsymbol{\Lambda}). \quad (7)$$

For our concrete case of a Gaussian copula, we first deduce the copula p.d.f. and then derive the ML estimate of $\boldsymbol{R}$. We know that the Gaussian copula is the copula of a multivariate Gaussian distribution with correlation matrix $\boldsymbol{R}$ and we know that the margins of the multivariate Gaussian are standard Gaussian distributions $\mathcal{N}(0,1)$. Hence, we just have to manipulate the p.d.f.

$$p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{R}) = \frac{1}{2\pi^{\frac{B}{2}} |\boldsymbol{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \boldsymbol{x}^T \boldsymbol{R}^{-1} \boldsymbol{x}\right) \quad (8)$$

such that we get an expression similar to Eq. (5). After some algebraic manipulations, it turns out that the p.d.f. of the Gaussian copula has the form

$$c(u_1,\ldots,u_B;\boldsymbol{R}) = |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\xi}^T(\boldsymbol{R}^{-1}-\mathbf{1})\boldsymbol{\xi}\right) \tag{9}$$

with $\boldsymbol{\xi} = [\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_B)]$. Then, it is straightforward to determine the ML estimate of $\boldsymbol{R}$ as

$$\hat{\boldsymbol{R}} = \sum_{i=1}^{M} \boldsymbol{\xi}_i^T \boldsymbol{\xi}_i \tag{10}$$

by taking the partial derivative w.r.t. $\boldsymbol{R}$ of the log-likelihood function corresponding to Eq. (9) and setting the resulting term to zero.

2.3. Similarity Measurement

For similarity measurement, we can directly employ the ML selection rule of the probabilistic CBIR framework, see Eq. (2). This is a natural choice, since it does not require to estimate any parameters of the unknown image's feature representation. We just have to perform the feature transformation step T and extract a collection of feature vectors $\boldsymbol{z}_1^*,\ldots,\boldsymbol{z}_K^*$ from the unknown image. We consciously use K instead of M to denote the number of feature vectors, since it is no requirement to use all vectors in the similarity measurement step. Eventually, we can write the ML selection rule as

$$g(\boldsymbol{z}_1^*,\ldots,\boldsymbol{z}_K^*) = \arg\max_j \sum_{i=1}^{K} \log p_{\mathbf{Z}}(\mathbf{z}_i^*;\boldsymbol{\Theta}^{(j)}). \tag{11}$$

The principle is visualized in Fig. 3. For the computation of Eq. (11), we have to evaluate the marginal p.d.f.s as well as the multivariate copula p.d.f. for each feature vector $\boldsymbol{z}_i^*$. Our choice of a Gaussian copula pays-off now, since evaluation of the Student t copula p.d.f. is a rather expensive operation in terms of computation time.

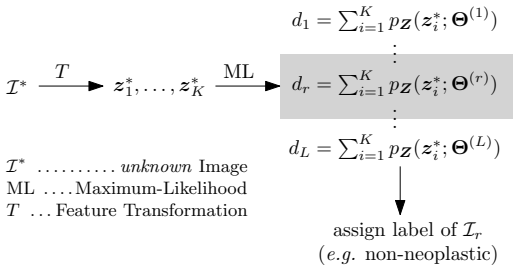


Figure 3: Illustration of the ML selection rule.

3. Experimental Study

Our original set of images consists of 269 RGB images (53 patients, either 624×533 or 586×502 pixel) acquired in 2005–2009 at the Department of Gastroenterology and Hepatology of the Medical University of Vienna using a zoom-endoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of $150\times$. All images were selected by the gastroenterologist conducting the colonoscopy with special emphasis to provide images with similar lightning conditions at approximately the same camera angle. To enhance the visual appearance of the mucosa, dye-spraying with indigo-carmine was applied and biopsies or mucosal resections were taken to obtain a histopathological diagnosis (*our ground truth*). The histology was obtained by a pathologist blinded to the colonoscopic procedure. Table 2 lists the distribution of Pit Patterns among all HMCC images. In order to increase the number of samples, we create an extended dataset by extracting subwindows of size 256×256 pixel from the original images such that the Pit Patterns are clearly distinctive and the subwindows contain a minimum number of specular reflections. This is a reasonable way to extend the dataset since the process closely resembles the clinical methodology. During colonoscopy, the gastroenterologist will typically look at more than just one region of an image. Eventually, the extended dataset contains 627 HMCC images.

I	II	III-S	III-L	IV	V
114	64	18	119	232	80

Table 2: Class distribution of the HMCC images.

3.1. Evaluation Setup

Our evaluation strategy is based on a *Leave-One-Out Cross Validation* (LOOCV) process [4]. Each image of the database is considered as an unknown image once and the class label is predicted by the system. The percentage of misclassified images represents the estimate of the prediction error of the whole system. Regarding to computation of the nearest neighbor, we differ to previously published works in one particular point. Up to now, the problem was considered from a purely classification oriented point of view. In such as setup, it does not matter which image is selected as the one to predict the class of an unknown image. The obtained classification rates, however, only convey an impression of how well an approach captures image information relevant for discrimination and are less interesting from the medical point of view. This becomes evident when we reconsider the fact that there is no restriction on the type of nearest neighbor, whatsoever. In the context of the dataset extension technique mentioned above, it is hence possible that the retrieved image stems from the same parent as the unknown image. In order to obtain clinically more meaningful classification rates, we modify the setup to resemble a more practical scenario. We impose the

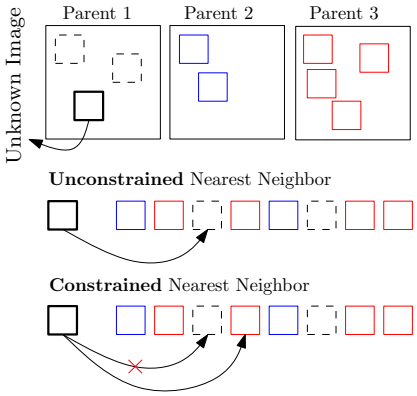


Figure 4: Illustration of the constrained and unconstrained nearest neighbor principle.

constraint that images are only admissible as nearest neighbors in case they do not stem from the same parent as the unknown image. We refer to this setup as the *constrained* setup, whereas the setup in previous works is referred to as the *unconstrained* setup. To visualize the difference, both setups are illustrated in Fig. 4.

3.2. Assessing Statistical Significance

In any reasonable comparative study on classification performance, situations occur where classification rates between two approaches seem very similar and do not allow to make any statements whether one approach performs better than the other. To have a quantitative measure of whether the class assignments of two approaches show significant differences, we employ a McNemar test [2]. Besides the 5x2 cross-validation test, this is one of the most popular and recommended tests for this purpose. The test statistic is based on counting the number of samples where approach A assigns the correct class label and approach B fails (denoted by n_{10}) and vice versa (denoted by n_{01}). Based on these counts the test statistic is defined as

$$S = \frac{(|n_{10} - n_{01}| - 1)^2}{n_{10} + n_{01}}. \tag{12}$$

In case the null-hypothesis of no statistically significant difference is true, S follows as Chi-Square distribution with one degree of freedom, *i.e.* $S \sim \chi^2_1$. In our comparative study, we plan to conduct n pairwise comparisons to the top approach at the 5% significance level. Such a multiple hypothesis testing scenario, however, requires to correct the significance level α. For that purpose, we employ two different strategies: (i) the classic Bonferroni correction to control the *Familywise Error Rate (FWER)*, *i.e.* $\alpha' = \alpha/n$ and (ii) we implement the *False-Discovery Rate (FDR)* control algorithm proposed by Benjamini & Hochberg [1].

3.3. Results

For a comparative study, we select three recently proposed discriminant classifier based approaches and two generative model based approaches which follow the principle of probabilistic CBIR. In the first group, we include the color-histogram approach of Häfner *et al*. [8], the Color-Eigen Subband features of Kwitt & Uhl [23] as well as the Weibull distribution features of Kwitt & Uhl [22]. In the second group, we include the CBIR approaches of Vasconcelos & Lippman [34] and Verdoolaege *et al*. [35]. Regarding the presentation of the results, we adhere to the convention to identify an approach by the names of the authors and the year of publication.

Tables 3 and 4 list the corresponding LOOCV results. In both classification setups, the copula approach exhibits the highest LOOCV rate. Besides, we notice that specificity is consistently higher than sensitivity in the two-class case. Hence, the diagnostic accuracy of neoplastic disease is higher than for non-neoplastic disease. In the three-class problem, the situation is reversed, signifying higher diagnostic accuracy of non-invasive neoplasia. We note, that the specificity is far superior than the 50% reported by Hurlstone *et al*. [10].

Finally, we take the copula-approach and perform pairwise comparisons to the other approaches to investigate if there is evidence for statistically significant differences in the classification results. A '**' next to the LOOCV accuracy of an approach signifies a significant difference using the corrected significance level based on either Bonferroni or Benjamini & Hochberg correction. As we see, the test results allow to report superior performance of our copula approach to all the other approaches in the two-class case. In the three-class case, only the CBIR method by Vasconcelos & Lippman [34] exhibits competitive performance. Compared to the currently highest reported LOOCV rates of $\approx 99\%$ in the two-class problem [7], we highlight that (i) these rates were obtained in the unconstrained nearest neighbor setup and (ii) the authors note that the approach might suffer overtraining issues due to feature subset selection [4]. Our approach, however, does not suffer any of these issues. In fact, overtraining is not possible at all, since no information about the class membership is used to estimate the feature representation.

4. Discussion

In this work, we presented a theoretically well-founded approach for the prediction of histological diagnosis of colorectal lesions. Motivated by the shortcomings of discriminant classifiers, we approached the problem from an image retrieval point of view. On the basis of a probabilistic formulation of CBIR, we introduced a novel, joint statistical model for complex wavelet coefficient magnitudes. The

Approach	Accuracy	Sensitivity	Specificity	PPV	NPV
Gaussian Copula, Weibull Margins	**96.65**	94.94	97.33	93.37	97.98
Vasconcelos & Lippman, 2000	94.74**	84.27	98.89	96.77	94.07
Verdoolaege *et al*., 2008	92.98**	91.01	93.76	85.26	96.34
Kwitt & Uhl, 2007	93.30**	91.01	94.21	86.17	96.36
Kwitt & Uhl, 2008	93.14**	90.45	94.21	86.10	96.14
Häfner *et al*., 2006	84.37**	74.16	88.42	71.74	89.62

Table 3: LOOCV results for the discrimination between non-neoplastic and neoplastic lesions.

Approach	Total	Non-Invasive vs. Invasive				
		Accuracy	Sensitivity	Specificity	PPV	NPV
Gaussian Copula, Weibull Margins	**93.46**	95.42	97.95	86.32	96.26	92.13
Vasconcelos & Lippman, 2000	92.50	97.28	98.26	93.81	98.26	93.81
Verdoolaege *et al*., 2008	88.52**	93.35	96.36	82.42	95.21	86.21
Kwitt & Uhl, 2007	89.47**	94.33	96.66	86.17	96.07	88.04
Kwitt & Uhl, 2008	88.84**	93.62	96.08	84.62	95.80	85.56
Häfner *et al*., 2006	78.31**	90.43	96.09	71.11	91.90	84.21

Table 4: LOOCV results for the discrimination between non-invasive and invasive neoplasia.

rule of Maximum-Likelihood image selection then proved to be a convenient way to predict the histology of an HMCC image. Our experiments show that we achieve significantly better prediction performance compared to recently proposed discriminant classifier approaches and slightly better performance than two state-of-the-art approaches from CBIR. In addition to that, the presented results for the non-neoplastic vs. neoplastic case are competitive to the results reported in medical literature and we achieve considerably better rates for the prediction of non-invasive vs. invasive neoplasia. We suggest, that this approach is not only beneficial in clinical practice, but also during the education of future gastroenterologists.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society – Series B*, 57(1):289–300, 1995. 6

[2] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, Oct. 1998. 6

[3] M. Do and M. Vetterli. Wavelet-based texture retrieval using Generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(2):146–158, Feb. 2002. 3

[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley & Sons, 2nd edition, Nov. 2000. 5, 6

[5] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray. Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colon mucosa. *IEEE Transactions on Information Technology in Biomedicine*, 2(3):197–203, Sept. 1998. 1

[6] K.-I. Fu, Y. Sano, S. Kato, T. Fuji, F. Nagashima, T. Yoshino, T. Okuno, S. Yoshida, and T. Fujimori. Chromoendoscopy using indigo carmine dye spraying with magnifying observation is the most reliable method for differential diagnosis between non-neoplastic and neoplastic colorectal lesions: a prospective study. *Endoscopy*, 36(12):1089–1093, Dec. 2004. 2

[7] M. Häfner, A. Gangl, R. Kwitt, A. Uhl, A. Vecsei, and F. Wrba. Improving pit-pattern classification of endoscopy images by a combination of experts. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '09)*, pages 247–254, London, UK, 2009. 6

[8] 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 J. R. Sveinsson, editor, *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG 2006)*, pages 58–61, Reykjavik, Iceland, June 2006. 6

[9] D. P. Hurlstone. High-resolution magnification chromoendoscopy: Common problems encountered in "pit pattern" interpretation and correct classification of flat colorectal lesions. *American Journal of Gastroenterology*, 97:1069–1070, 2002. 2

[10] D. P. Hurlstone, S. S. Cross, I. Adam, A. J. Shorthouse, S. Brown, D. S. Sanders, and A. J. Lobo. Efficacy of high magnification chromoscopic colonoscopy for the diagnosis of neoplasia in flat and depressed lesions of the colorectum: a prospective analysis. *Gut*, 53(2):284–290, February 2004. 1, 2, 6

[11] 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, October 2006. 1

[12] 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, 2005 (CBMS'05)*, pages 575–580, Dublin, Ireland, June 2005. 1

[13] H. Joe. *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1997. 4

[14] S. Karkanis. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, Sept. 2003. 1

[15] 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, 2001 (ICIP'01)*, pages 833–836, Thessaloniki, Greece, October 2001. 1

[16] 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: WJG*, 12(9):1416–1420, March 2006. 2

[17] S. Kato, T. Fujii, I. Koba, Y. Sano, K. Fu, A. Parra-Blanco, H. Tajiri, S. Yoshida, and B. Rembacken. Assessment of colorectal lesions using magnifiying colonoscopy and mucosal dye spraying: Can significant lesions be distinguished? *Endoscopy*, 33:306–310, Apr. 2001. 2

[18] K. Konishi, K. Kaneko, T. Kurahashi, T. Yamamoto, M. Kushima, A. Kanda, H. Tajiri, and K. Mitamura. A comparison of magnifying and nonmagnifying colonoscopy for diagnosis of colorectal polyps: a prospective study. *Gastrointestinal Endoscopy*, 57:48–53, 2003. 2

[19] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Chapman & Hall, 2006. 4

[20] 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, 1998 (EMBS'98)*, Hong Kong, China, October 1998. 1

[21] 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, 1994. 1, 2

[22] 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, 2007. 3, 6

[23] 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, 2008. 6

[24] R. Kwitt and A. Uhl. Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval. In *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, pages 933–936, San Diego, California, United States, Oct. 2008. 3

[25] R. Kwitt and A. Uhl. A joint model of complex wavelet coefficients for texture retrieval. In *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pages 1877–1880, Cairo, Egypt, Nov. 2009. 4

[26] 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, February 2003. 1

[27] A. Meining, T. Rösch, R. Kiesslich, M. Muders, F. Sax, and W. Heldwein. Inter- and intra-observer variability of magnification chromoendoscopy for detecting specialized intestinal metaplasis at the gastroesophageal junction. *Endoscopy*, 36(2):160–164, Feb. 2004. 2

[28] B. M. Namee, P. Cunningham, S. Byrne, and O. Corrigan. The problem of bias in training data in regression problems in medical decision support. *Artificial Intelligence in Medicine*, 24(1):51–70, Jan. 2001. 3

[29] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, second edition, 2006. 4

[30] 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. 3

[31] J. K. Shuttleworth, A. G. Todman, R. N. G. Naguib, B. M. Newman, and M. K. Bennett. Colour texture analysis using Co–Occurrence matrices for classification of colon cancer images. In *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE'02)*, volume 2, pages 1134–1139, Winnipeg, Manitoba, Canada, 2002. 1

[32] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institute de Statistique de l'Université de Paris*, 8:229–231, 1959. 4

[33] S.-Y. Tung, C.-S. Wu, and M.-Y. Su. Magnifying colonoscopy in differentiting neoplastic from nonneoplastic colorectal lesions. *American Journal of Gastroenterology*, 96:2628–2632, 2001. 2

[34] N. Vasconcelos and A. Lippman. A probabilistic architecture for content-based image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'00)*, pages 1216–1221, Hilton Head, South Carolina, United States, June 2000. 3, 6

[35] G. Verdoolaege and P. S. S. De Backer. Multiscale colour texture retrieval using the geodesic distance between multivariate Generalized Gaussian models. In *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, pages 169 – 172, San Diego, California, USA, Oct. 2008. 6