

Andreas Vécsei, Thomas Fuhrmann, Michael Liedlgruber, Leonhard Brunauer, Hannes Payer and Andreas Uhl, “Automated classification of duodenal imagery in celiac disease using evolved Fourier feature vectors ”, *Computer Methods and Programs in Biomedicine*, 95(Suppl. 2), pp. S68-S78 (2009).

doi:10.1016/j.cmpb.2009.02.017

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

Andreas Vécsei², Thomas Fuhrmann¹, Michael Liedlgruber¹, Leonhard Brunauer¹,
Hannes Payer¹, and Andreas Uhl^{1,*}

¹Department of Computer Sciences, Salzburg University, Austria
²St. Anna Children's Hospital, Vienna, Austria

*Corresponding author e-mail: uhl@cosy.sbg.ac.at

Abstract

Feature extraction techniques based on selection of highly discriminant Fourier filters have been developed for an automated classification of magnifying endoscope images with respect to pit patterns of colon lesions. These are applied to duodenal imagery for diagnosis of celiac disease. Features are extracted from the Fourier domain by selecting the most discriminant features using an evolutionary algorithm. Subsequent classification is performed with various standard algorithms (KNN, SVM, Bayes classifier) and combination of several Fourier filters and classifiers which is called multiclassifier. The obtained results are promising, due to a high specificity for the detection of mucosal damage typical of untreated celiac disease.

1. Introduction

Celiac disease is a complex autoimmune disorder that affects the small bowel in genetically predisposed individuals of all age groups after introduction of gluten containing food. Commonly known as gluten intolerance, this disease has several other names in literature, including cœliac disease, c(o)eliac sprue, non-tropical sprue, endemic sprue, gluten enteropathy or gluten-sensitive enteropathy. Characteristic for the disease is an inflammatory reaction in the mucosa of the small intestine caused by a dysregulated immune response triggered by ingested gluten proteins of certain cereals (wheat, rye, and barley), especially against gliadine. During the course of the disease the mucosa loses 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. Besides standard upper endoscopy, several new endoscopic approaches for diagnosing celiac disease have been applied.

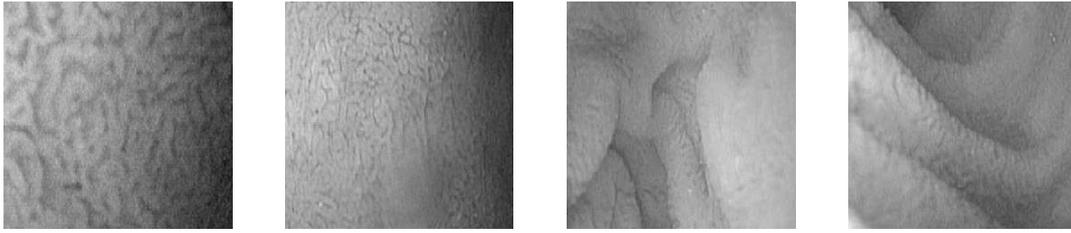


Figure 1. Negative celiac disease indication: normal mucosal state.

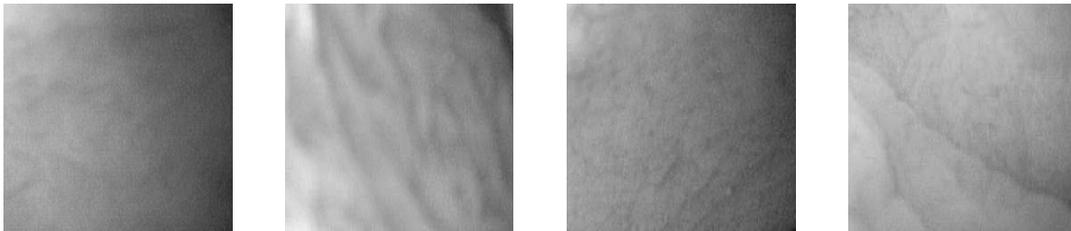


Figure 2. Positive celiac disease indication: villous atrophy.

The modified immersion technique described in [1] is based on the instillation of water into the duodenal lumen for better visualization of the villi. Furthermore magnifying endoscopy (standard endoscopy with additional magnification) has been investigated [2]. For the performance of capsule endoscopy [3] the patient swallows a small capsule equipped with a camera that takes images of the duodenal mucosa during its passage through the intestine. All these techniques aim for detection of total or partial villous atrophy and other specific markers. These specific markers show a high specificity for celiac disease in adult patients if all of them are found during endoscopy: scalloping of the small bowel folds, reduction in the number or loss of Kerkring's folds, scalloped folds, mosaic patterns, and visualization of the underlying blood vessels [4].

Figures 1 and 2 show examples for duodenal images with positive and negative indications for celiac disease.

During endoscopy at least three duodenal biopsies are taken. Microscopic changes within these specimen are classified according to Marsh classification [5].

The whole procedure of celiac disease diagnosis, including duodenoscopy with biopsy, is a time-consuming and cost-intensive process. To save costs, time, and manpower and simultaneously increase the safety of the procedure it would be desirable to develop an approach avoiding biopsy. Recent studies [1, 6] investigating such endoscopic techniques report reliable results. These could be further improved by analysis of the acquired visual data (digital images and video sequences) with the assistance of computers.

In this work we apply several techniques that were developed for Pit pattern classification of magnifying endoscopic images to imagery obtained from small bowel endoscopy. The aim is the detection of the villous atrophy and classification with respect to its extent (no villous atrophy, partial, or total villous atrophy).

2. Automated Classification

Computer based image classification is usually performed in three steps. The first step is to detect image regions that show high informational content for the particular classification problem. Since almost no classification algorithms can cope with high-dimensional input data, we have to reduce the dimension of these image regions. Next, the discriminative information of an image is encoded by a numerical feature vector describing the relevant information. Various techniques can be used for this process. Finally, the features are used for training of the classification algorithm (i.e. determining the optimal parameters for the specific classification problem). Any unknown image that is presented to the classification algorithm is then classified according to the previously learned settings.

2.1 Image preprocessing

Many of the images used throughout our experiments are very blurry. Apart from that they often suffer from a low contrast. This is why we decided to carry out experiments with preprocessing too.

To enhance the contrast we use an advanced contrast enhancement technique called CLAHE (Contrast Limited Adaptive Histogram Equalization) [8]. Compared to other contrast enhancement algorithms (e.g. histogram equalization), this algorithm operates on local image regions. For this purpose the image is subdivided into image tiles (so-called contextual regions) and the contrast is enhanced within each of these regions. To avoid artifacts between two adjacent tiles an interpolation algorithm is employed. Apart from that, to avoid amplification of present noise, the contrast enhancement is limited within homogenous regions (which can be identified easily by high peaks in the histogram of the according region).

To sharpen the images we use Laplace sharpening [9]. Basically this technique computes a gradient image using convolution with a suitable kernel. This gradient image is then added to the original image. Usually this algorithm is used with a rather small kernel (e.g. 3x3). Regarding the images available this would very often result in highlighting present noise, which is the main reason for using a 9x9 kernel throughout our experiments. The effect of the preprocessing steps is shown in Figure 3.

2.2 Fourier-based Feature Extraction

For classification we employ features from the Fourier domain [10]. Multiple ring shaped filters with variable widths are applied to the centered Fourier spectrum of each color channel (RGB) for selecting relevant subsets of the most discriminative coefficients that keep the scatter within each class small and give a high interclass variability. Statistical measures for the coefficients of each ring (mean, standard deviation) compose the feature vector. The main challenge with this method is finding the optimal ring configuration (i.e. the number of rings which determines the length of the vector and the width of each ring).

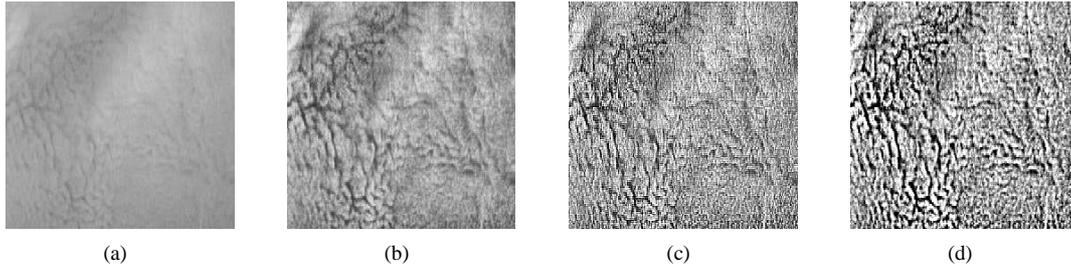


Figure 3. The effect of the preprocessing steps. (a) original image (b) CLAHE applied (c) CLAHE with subsequent sharpening applied (using a 3x3 kernel) (d) same as (c), but using a 9x9 sharpening kernel.

Parameter	Value
Bit mutation rate p_m	$5/l_c$
Number of generations g	100
Crossover points	2
Crossover rate p_c	0.6
Population size n	500

Table 1. Parameters used for the evolutionary search of the ring filters.

For this problem evolutionary optimization methods tend to provide good results [11]. Having a dynamic number of rings with a dynamic amount of Fourier coefficients (i.e. width of the ring), we try to minimize the classification error rate.

For this purpose the JEvolution [12] framework has been used, which is a lean and compact Java framework for Evolutionary Algorithms supporting standard EA components (e.g. different genotype encodings, common mutation and recombination operators) and an interface for problem-specific code (i.e. the fitness evaluation).

As we are interested in the number of rings and width of each ring, the ring filter setting is encoded in a bit-chromosome. The first 4 bits are used to encode the total number of rings (max. 15 rings). Additionally there are 5 bits used to encode the width of each of the 15 rings (1 sign bit for the sign and 4 bits for the width). This results in a total chromosome length l_c of 79 bits. Rings having a negative width represent a gap of the given width. It is important to mention that the ring widths are relative to the image width. For example, if the encoding consists of two rings of width 1 and 3 (sum is 4) and the image width equals 256 pixel (half image width is 128), then the first ring is 32 (128/4) pixels wide and the second ring is 96 (3*128/4) pixels wide.

The parameters used to configure the JEvolution framework are summarized in Table 1. Selection is realized by binary tournament selection without replacement. The fitness of an individual is the number of correctly classified images determined by a leave-one-out cross validation over the total image set [13]. While ring configurations showing a high fitness value (low classification error rates) are preserved in the evolutionary process, chromosomes with a low fitness value are dropped.

In order to trim the classification algorithm to favor the correct classification of one of the classes (e.g. the clinically more relevant class showing villous atrophy) we also employ fitness functions that minimize the error rate of one single class instead of optimizing the overall classification rate which can have low significance for unbalanced data sets.

In [11] it is shown that feature vectors including information from all color channels usually give better results. We follow this strategy and extract feature vectors for each distinct color channel.

2.3. Classification

For feature vector classification we employ three common machine learning algorithms, namely the k-Nearest Neighbor classifier (k-NN), Support Vector Machines (SVM), and a statistical Bayes classifier.

The k-NN classifier [13] is one of the simplest classification algorithms. Classification is done by finding the k closest neighbors of an input feature vector x in the feature space according to some distance metric. The unknown input vector x is then assigned to the class to which the majority of the k nearest neighbors belongs to.

The SVM classifier, further described in [14], aims for constructing classifying hyperplanes which are optimal for the separation of the given feature vectors.

Bayes classification [15] is based on applying Bayes' theorem on an independent feature model and requires only a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

2.4. Multiclassification

Although we achieve good results with the classifiers mentioned above (c.f. Section 3.2) we additionally combine different Fourier filters and classifiers into a multiclassifier that is based on an reliability measure for each ring configuration contained in the multiclassifier.

The idea behind the reliability measure is to determine how reliable a method is. That is, how reliable is a given classifier with respect to the resulting image class for a given image. To compute this measure we employ the Bayesian probability and get the reliability A_m for each filter/classifier combination m by using equation (1).

$$A_m = \left(\sum_{i=1}^C N_i \frac{b_i r_i}{(1 - b_i)(1 - r_i) + b_i r_i} \right) / \frac{N}{2} - 1 \quad (1)$$

C denotes the number of image classes used, N_i denotes the number of images in class i , N denotes the total number of images used, b_i is the probability that an image is in class i and r_i is the classification rate for class i . Factoring the inner part of the sum by N_i results in a weighted reliability measurement. This is necessary since the number of images is unbalanced across the image classes. Finally, the sum is divided by $N/2$ and decremented by 1 to map the reliability to a value between -1 and 1.

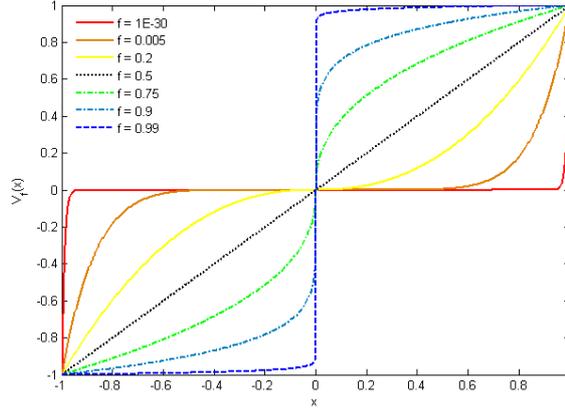


Figure 4. The effect of different choices for f in the value remapping function.

To be able to control how strong the method reliability influences the result of the multiclassifier, the value is remapped by using the following equation:

$$V_f(x) = \text{sign}(x)|x|^{\ln(f)/\ln(0.5)} \quad (2)$$

where x is the value to be remapped and f is the parameter which controls the shape of the remapping curve.

The effect of choosing different values for f on the shape of the remapping curve is depicted in Figure 4. As we can see in this figure, choosing $f = 0.5$ corresponds to a linear mapping (i.e. the mapping has no effect at all).

The resulting image class c_i for an image i is computed by using the following equation:

$$c_i = \sum_{j=1}^M D_{i,j} V_f(A_j) \quad (3)$$

where M is the number of combinations of filters and classifiers used, $D_{i,j}$ is the remapped class number, which has previously been assigned to image i by filter/classifier combination j (-1 for class 0 and 1 for class 1), and A_j is the reliability of filter/classifier combination j . If $c_i < 0$ the resulting class is assumed to be 0. For $c_i > 0$ the image is labeled as class 1.

3. Experimental Study

3.1. Settings

Experiments are conducted with an image database consisting of images from two regions of the small intestine referred to as *Bulbus duodeni* and *Pars descendens*. These images were taken at the St. Anna's Children Hospital in Vienna during 2006 and 2007 using a standard duodenoscope without magnification. Subsequent biopsy and histological examination of the mucosal state (according to Marsh classification) has led to the final results for the test database shown in Table 2. *Non-Celiac*

	No Celiac	Celiac	Total
Bulbus	162	30	192
Pars	150	49	199

Table 2. Number of images per class.

denotes visibility of villous structures, while *Celiac* indicates that no villi could be detected.

Due to the rather limited number of images available for our experiments we employ leave-one-out cross validation [13]. That is, one image is picked from the database to form the validation data, while the remaining images are used to train the respective classifier. This process is repeated for each image in the database ending up with the overall classification rate.

3.2. Results with standard classifiers

Tables 3 and 4 show the results we obtained using the evolutionary algorithm to find the optimal ring configurations for the respective classifiers. R_T denotes the overall classification result. R_0 , as defined in equation (4), indicates the method's specificity (True Negative Rate) which is the percentage of correctly classified images actually showing a normal mucosal state. R_1 , as defined in equation (5), refers to the sensitivity (True Positive Rate) which is the percentage of correctly classified images showing villous atrophy

$$R_0 = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (4)$$

$$R_1 = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (5)$$

As already pointed out above, the image database is rather unbalanced with respect to the number of images in the image classes. Thus we decided to carry out the evolutionary tests using three different types of fitness criteria, denoted by C_T , C_1 , and C_S . The fitness criterion C_T is the total classification rate, C_1 is the classification rate for the Celiac class (i.e. the sensitivity), and C_S is the sum of the classification rates over all image classes.

Since we needed to find an optimal k -value for the k-NN classifier, we decided to encode the k -value into an additional bit chromosome (when using the k-NN classifier). However, it turned out that the optimal value for k is either 1 or 2 for all cases listed in tables 3 and 4. In case of the SVM classifier we tested to evolve the optimal SVM parameters too, but a full grid search always yielded better overall classification results.

According to Table 3 the Bayes classifier delivers the best overall classification result for the Bulbus images of 94%. Apart from that we can see, that no matter which fitness criterion we use, the results remain the same for each of the classifiers. The worst overall classification results of 90% and 91% have been obtained using the k-NN classifier due to low sensitivity. The

	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1
	C_T			C_1/C_T			C_1/C_S			C_S		
Bulbus												
k-NN	90	99	40	91	99	50	85	91	60	90	98	47
SVM	93	96	77	92	96	70	86	88	88	93	98	63
Bayes	94	97	77	93	100	57	92	93	83	94	99	63
Pars Descendens												
k-NN	84	96	47	83	97	41	81	87	65	84	89	71
SVM	86	94	61	82	89	61	81	84	73	84	92	59
Bayes	85	91	65	80	89	51	73	72	78	86	95	59

Table 3. Detailed results for the single filter configurations obtained by using an evolutionary algorithm.

	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1
	C_T			C_1/C_T			C_1/C_S			C_S		
Bulbus												
k-NN	92	99	53	92	99	50	91	96	63	92	99	50
SVM	93	99	60	93	96	80	93	96	80	92	98	60
Bayes	97	99	83	91	98	50	86	86	90	93	98	67
Pars Descendens												
k-NN	86	91	73	85	99	45	85	89	73	85	97	49
SVM	86	96	57	84	93	57	79	83	69	85	89	76
Bayes	88	99	55	79	90	47	71	68	80	86	95	61

Table 4. Detailed results for the single filter configurations obtained by using a genetic algorithm with mutation disabled ($mr = 0$).

column C_1/C_T denotes the best results obtained with respect to the overall classification rate using the fitness criterion C_1 . Similarly, C_1/C_S denotes the best results obtained with respect to the sum of the specificity and sensitivity when using the fitness criterion C_1 .

For the Pars Descendens images the SVM classifier and the Bayes classifier yield the best overall classification rate of 86% when using the fitness criterion C_T and C_S , respectively. The k-NN classifier again delivers the worst overall classification result of 83% when using fitness criterion C_1 . Again, changing the fitness criterion has no great impact on the classification results. Only the results for the Bayes classifier drop significantly by at least 5% when using criterion C_1 .

To investigate whether the results can be improved when using no genetic mutation of the bit chromosomes at all, we also carried out evolutionary tests with the mutation rate set to 0. The results of these tests are shown in Table 4. Interestingly in most cases the classification results have indeed been improved by disabling mutation. When using the SVM classifier the results remain roughly the same with variations of only 1%. In the case of the k-NN classifier an improvement of 1-2% can be observed. The highest improvement of 3% is achieved when using the Bayes classifier with criterion C_T . However, when using criterion C_1 or C_S the results remain the same or even drop slightly by 1-2%.

	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1
	C_T			C_1/C_T			C_1/C_S			C_S		
Bulbus												
k-NN	91	99	50	92	99	50	88	91	70	92	99	50
Bayes	93	98	63	92	96	70	89	91	87	93	98	70
Pars Descendens												
k-NN	85	99	43	85	99	41	83	89	65	84	99	41
Bayes	86	95	59	86	95	57	79	81	76	86	96	55

Table 5. Detailed results for the single filter configurations obtained by using an evolutionary algorithm (with CLAHE and sharpening applied).

	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1	R_T	R_0	R_1
	C_T			C_1/C_T			C_1/C_S			C_S		
Bulbus												
k-NN	92	96	70	93	99	63	90	92	77	93	99	63
Bayes	97	100	80	92	99	53	85	85	87	95	96	87
Pars Descendens												
k-NN	85	92	65	86	91	69	86	91	69	86	92	67
Bayes	93	99	73	80	89	55	80	82	76	91	95	82

Table 6. Detailed results for the single filter configurations obtained by using a genetic algorithm with mutation disabled ($mr = 0$) and preprocessing applied (CLAHE and sharpening).

Figures 5 and 6 show the different ring configurations used to obtain the classification results for the single tests using Bulbus images and Pars Descendens images, respectively. Each triplet of images represents the ring filters used to obtain the features for the R, G, and B channel of the RGB color model, respectively. From the filters shown in these figures we can see that there exist many very different ring configurations for filters which yield good results.

3.3 Results with preprocessing applied

As already pointed out in Section 2.1, we also carried out experiments using preprocessed images. The goal of these experiments was to improve the classification results (especially the sensitivity values). As we can see from Table 5, the results remain roughly the same when mutation is enabled. This applies for the Bulbus images as well as for the Pars Descendens images. All in all the values R_T differ only slightly with variations between -3% and +3%. Only in some cases we were able to achieve an improvement regarding R_T as well as a major improvement regarding R_1 (the sensitivity). These cases are shown in bold in Table 5.

In the case of the Bulbus images and the k-NN classifier R_T has been improved by 1% and the respective value of R_1 has been increased by 10%. When looking at the best k-NN result with respect to the fitness criterion C_S the improvement of R_T is 3% while R_1 is increased by 10%. In both cases the specificity remains equal to the non-preprocessing results. In the case

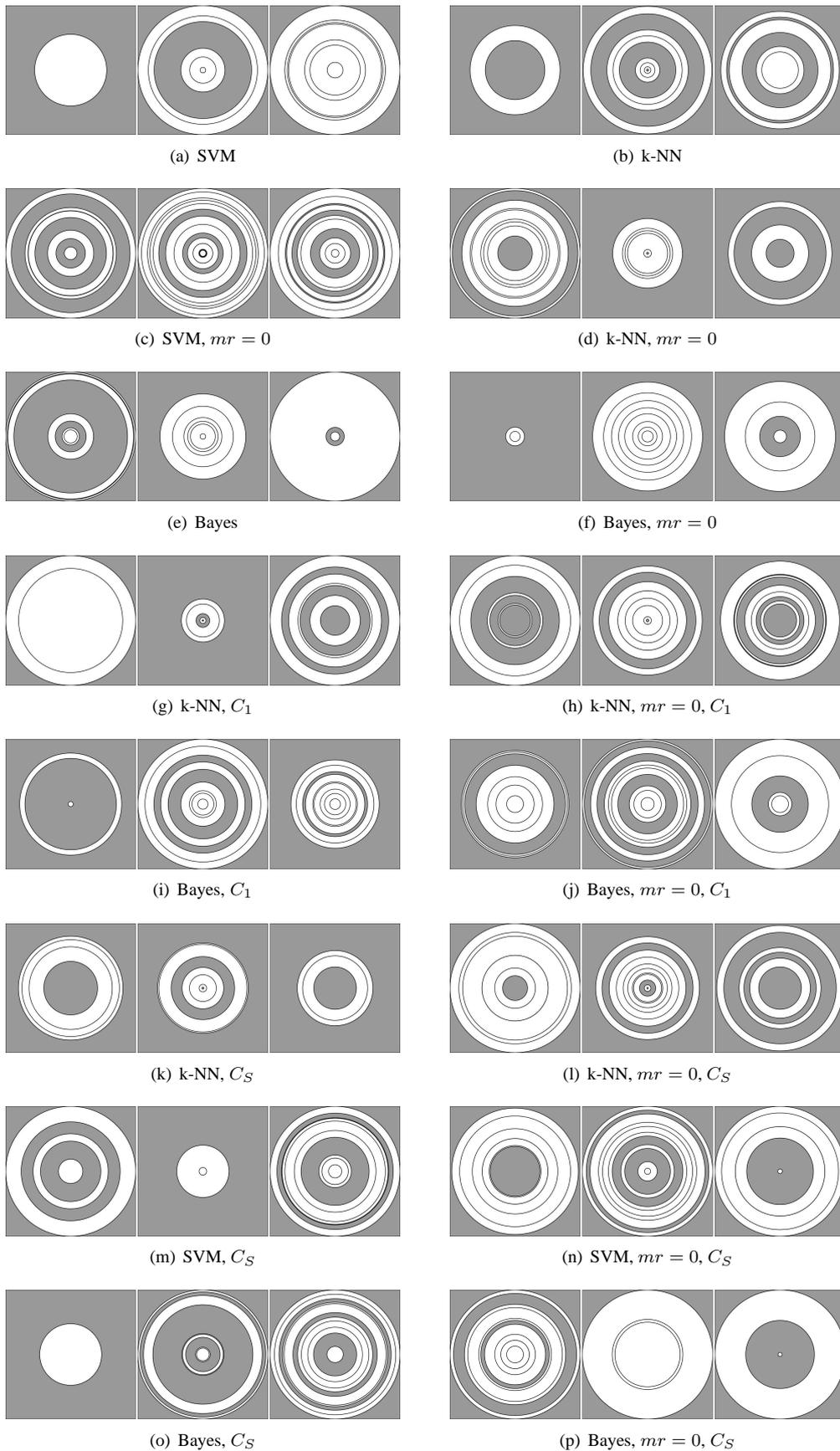


Figure 5. The filters used to obtain the fittest individuals for the single tests (Bulbus) where each triple represents the the R, G and B channel (from left to right).

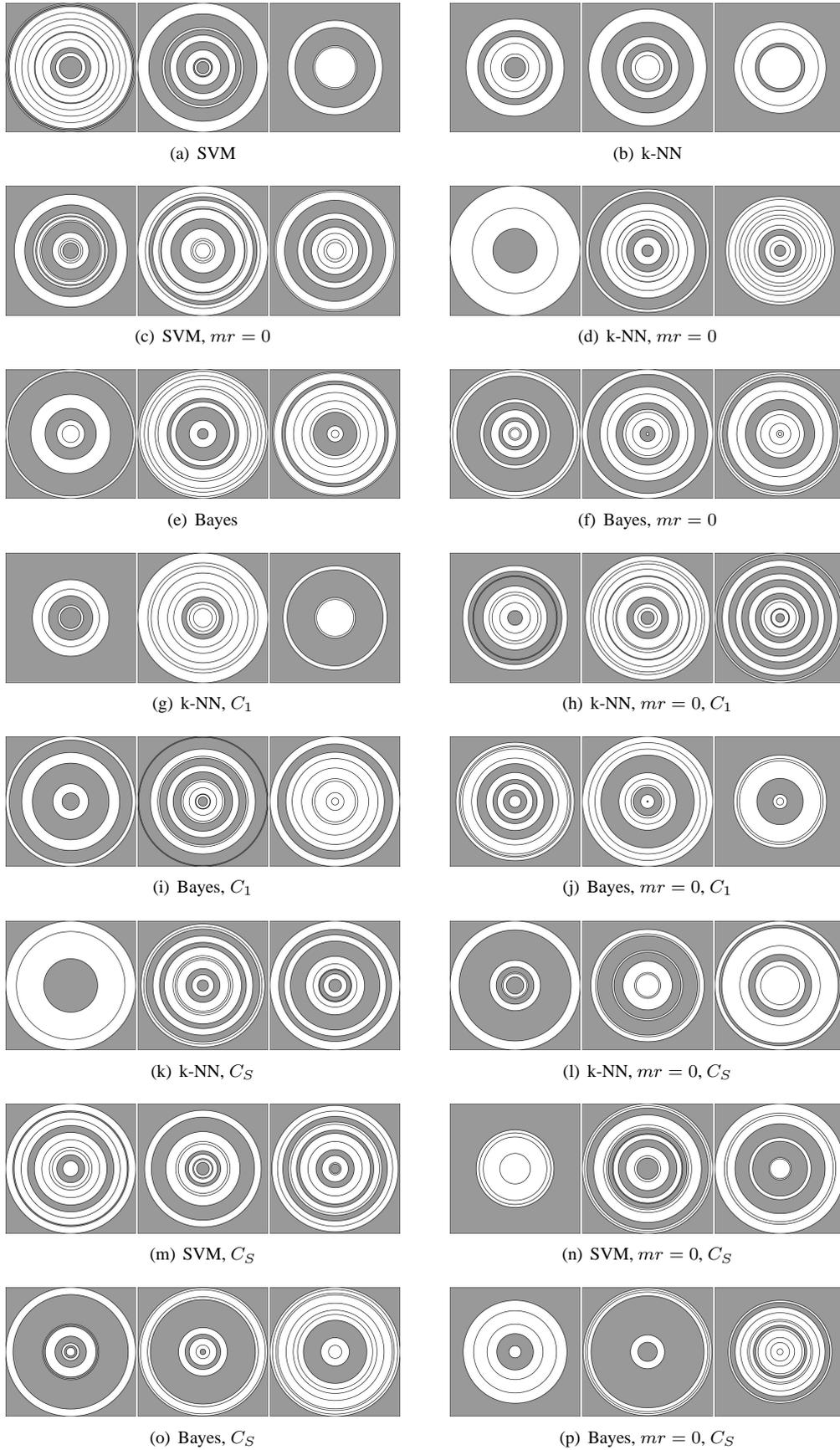


Figure 6. The filters used to obtain the fittest individuals for the single tests (Pars Descendens) where each triple represents the the R, G and B channel (from left to right).

of the Pars Descendens images only the Bayes classifier with respect to criterion C_T was able to deliver an improvement for R_T and R_1 by 6%.

Regarding the experiments with mutation disabled, the picture is a little bit different, as we can see from Table 6. While still most variations of R_T lie between -1% and +2%, there are more cases in which R_T and R_1 show better results simultaneously. Again, these cases are shown in bold in Table 6.

In the case of the Bulbus images and the k-NN classifier the best result with respect to C_T has been improved by 1% while the respective sensitivity has been increased by 13%. The same changes apply for the case when criterion C_S is used. The Bayes classifier was able to deliver an increase for R_T of 2% and an improvement of 20% for the respective sensitivity. Regarding the Pars Descendens images, the k-NN classifier again delivers better results with respect to C_T and C_S . While R_T is increased by 1% in both cases, the sensitivity has been improved by 24% and 18%, respectively. When using the Bayes classifier we see a notable improvement of R_T for criterion C_T , as well as for criterion C_S , of 5%. The respective sensitivities have been increased by 25% and 21%, respectively.

All in all we observe rather low sensitivity values (compared to the overall classification rate and the specificity). One possible explanation for this behavior is the rather unbalanced image set used throughout our experiments. As a consequence, regarding the first image class (containing images showing villous structures), the classifiers have much more samples to learn their parameters from.

3.4. Results with multiclassification

From tables 3 and 4 we see that each method's sensitivity is always considerably lower than its specificity. Therefore we combined several single filter configurations found by different evolutionary configurations using different classifiers to get a multiclassifier as described in Section 2.4.

The aim of this combination is to get higher classification results, especially higher sensitivities. But since there exists no clear way to determine the configurations which should be combined for the multiclassifier, the ring filters have been found in a heuristic way. For this purpose we tested several different combinations of configurations which have been found using the different fitness criteria and which yielded good results for one of the image classes or the overall classification result. The combinations shown in Table 7 turned out to be optimal with respect to the overall classification rate of the multiclassifier. As we can see, the sensitivities of the chosen filters used for the combination for the Bulbus images are high compared to the sensitivities shown in tables 3 and 4.

As we can see from Table 8, the results obtained with the combination of different ring filters are indeed better. Compared to the results in tables 3 and 4 we have an improvement of 1% in terms of the overall classification rate for Bulbus images and an improvement of 3% for Pars Descendens images. While in the latter case we have an improvement of 3% compared to the Bayes result in Table 4, the improvement is not that high for the Bulbus images (1% for the Bayes result in Table 4).

	R_T	R_0	R_1
Bulbus			
SVM, C_T	93	96	77
Bayes, C_T	94	97	77
Bayes, C_T , mutation disabled	97	99	83
Pars descendens			
k-NN, C_T , mutation disabled	83	99	35
k-NN, C_T , mutation disabled	86	91	73
Bayes, C_T	85	91	65
Bayes, C_T , mutation disabled	88	99	55
k-NN, C_1 , mutation disabled	85	99	45
k-NN, C_S	84	89	71

Table 7. The different configurations combined into the multiclassifier.

	R_T	R_0	R_1
Bulbus	98	100	87
Pars Descendens	91	99	65

Table 8. The percentage of correctly classified images obtained when using the combination of several different ring filters.

Considering sensitivity in the Bulbus case, we have an improvement of at least 4%. However, the filter combination used for the Pars Descendens images is outperformed by some of the single configurations in terms of sensitivity. The combination for the Bulbus images consists of three distinct ring filters. The first two ring configurations have been found by carrying out two evolutionary searches using the Bayes classifier with disabled ($mr = 0$) and enabled mutation, respectively. The third configuration is the result of an evolutionary search using the SVM classifier with a full grid search enabled for the SVM parameters. The ring filters according to the configurations found are shown in figures 7(a)- (c).

The combination for the Pars Descendens images consist of six different ring filters. The first two ring configurations have

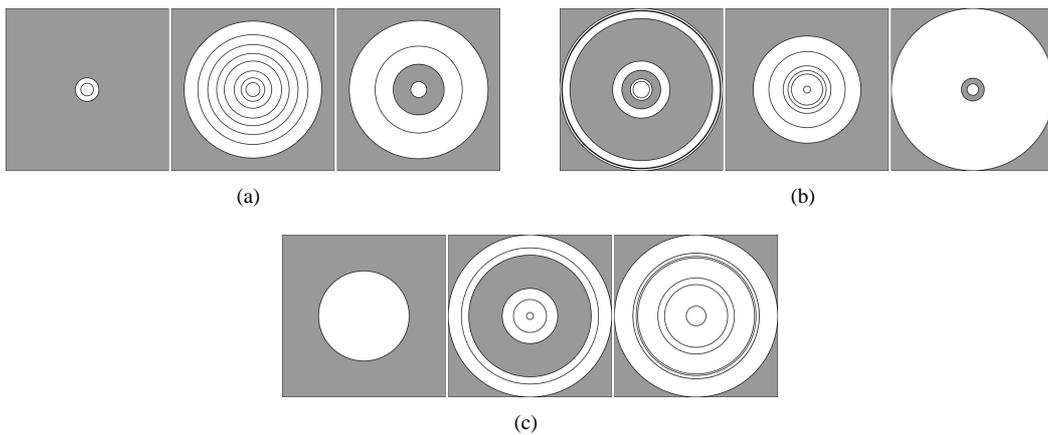


Figure 7. The filters used for the multiclassifier for the Bulbus images.

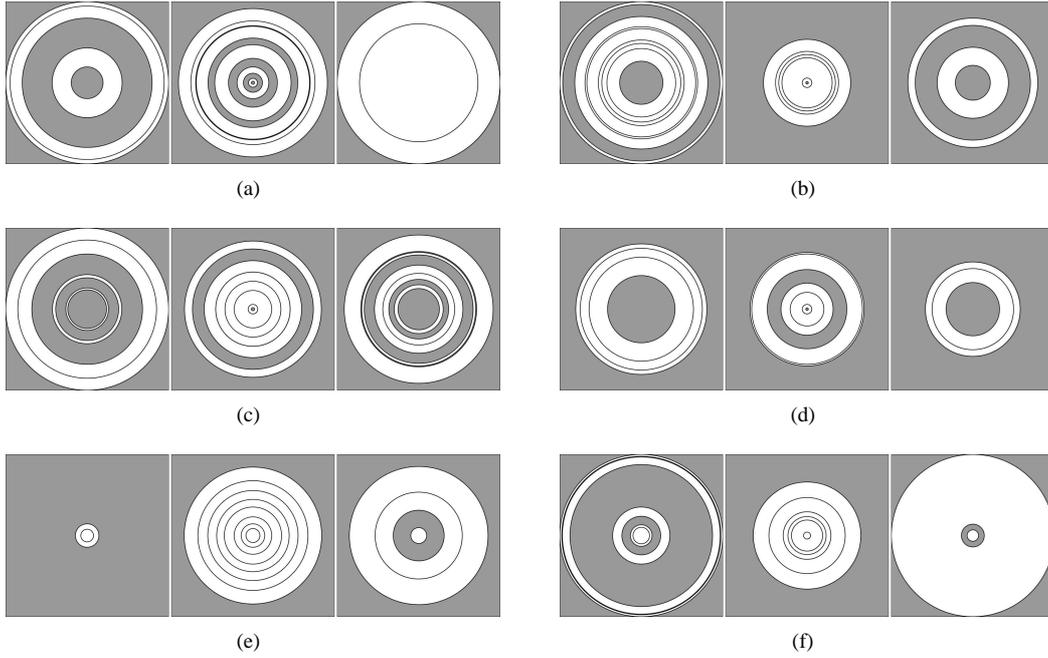


Figure 8. The filters used for the multiclassifier for the Pars Descendens images.

	R_T	R_0	R_1
Bulbus	97	99	83
Pars Descendens	94	99	78

Table 9. The percentage of correctly classified images obtained when using the combination of several different ring filters (with preprocessing).

been found by performing an evolutionary search in conjunction with the k-NN classifier and a disabled mutation. The third ring configuration is the result of the k-NN classifier with disabled mutation too, but using the fitness criterion C_1 . The fourth ring configuration is also the result of the k-NN classifier but with enabled mutation and using the fitness criterion C_S . The last two ring configurations are the result of using the Bayes classifier for the evolutionary search with disabled and enabled mutation, respectively. The ring filters according to the configurations found are depicted in figures 8(a)-(f).

3.5 Result of multiclassification with preprocessing applied

Based on the experiments carried out with preprocessing applied to the images, we also tested to improve the multiclassifier results presented above. Again, we heuristically determined the optimal combination of ring filters (from the configurations shown in tables 5 and 6) since there is no clear way on how to choose the best combination.

The new results of the multiclassifier are shown in Table 9. Regarding the Bulbus images we see that the results are lower compared to the results without preprocessing. While the specificity dropped by 1% only, the sensitivity dropped by

	R_T	R_0	R_1
Bulbus			
Bayes, C_T , mutation disabled	97	100	80
Bayes, C_S	93	98	70
Bayes, C_S , mutation disabled	95	96	87
Pars descendens			
Bayes, C_T , mutation disabled	93	99	73
k-NN, C_1 , mutation disabled	86	91	69
Bayes, C_S , mutation disabled	91	95	82

Table 10. The different configurations combined into the multiclassifier using preprocessing.

4%. However, at least in the case of the Pars images we were able to improve the overall classification rate as well as the sensitivity by 3% and 13%, respectively.

The configurations combined to obtain the new results are shown in Table 10. In the case of the Bulbus images we combined three configurations which are all based on the Bayes classifier using criteria C_S and C_T , two times with a disabled mutation. For the Pars images we again combined three different configurations with a disabled mutation, two times based on the Bayes classifier using criteria C_T and C_S and once based on the k-NN classifier using the criterion C_1 . Note that Table 10 also reveals a possible reason for the weaker performance of the multiclassifier when preprocessing is applied: No SVM configurations have been included contrasting to the case without preprocessing. When including such cases, we expect to improve the multiclassifier results as well as it is the case for single k-NN and Bayes classifiers.

4. Conclusion

Our results have shown that the detection of villous atrophy in duodenal images by an automated classification using evolved Fourier feature vectors is feasible. By combining the results of several filter/classifier methods we can further improve both specificity and sensitivity. We have also shown that we are able to get better results in some cases by applying preprocessing. This partially applies to the multiclassifier too, as we have seen in the case of the Pars images.

Sensitivity consistently shows lower rates as compared to specificity - images exhibiting villous atrophy in certain regions may also consist of regions without any visible degradation which results in a much more difficult classification task. In future work we will focus on a more localized classification, so that images containing only some regions with villous atrophy are potentially correctly classified as well.

5. Acknowledgments

This work has been supported by the Austrian National Bank "Jubiläumsfonds", project no. 12991.

References

- [1] G. Cammarota, P. Cesaro, A. Martino, et al. High accuracy and cost-effectiveness of a biopsy-avoiding endoscopic approach in diagnosing coeliac disease. *Alimentary Pharmacology and Therapeutics*, 23(1):61–69, January 2006.
- [2] G. Cammarota, A. Martino, and G. A. Pirozzi. Direct visualization of intestinal villi by high-resolution magnifying upper endoscopy: a validation study. *Gastrointestinal Endoscopy*, 60(5):732–738, 2004.
- [3] R. Petroniene, E. Dubcenco, and J. P. Baker. Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. *The American Journal of Gastroenterology*, 100(3):685–694, March 2005.
- [4] S. Niveloni, A. Florini, R. Dezi, et al. Usefulness of videoduodenoscopy and vital dye staining as indicators of mucosal atrophy of celiac disease: assessment of interobserver agreement. *Gastrointestinal Endoscopy*, 47(3):223–229, March 1998.
- [5] 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, 1992.
- [6] G. Cammarota, L. Cuoco, P. Cesaro, et al. A highly accurate method for monitoring histological recovery in patients with celiac disease on a gluten-free diet using an endoscopic approach that avoids the need for biopsy: a double-center study. *Endoscopy 2007*, 39(1):46–51, January 2007.
- [7] N. Chand and A. A. Mihas. Celiac disease: Current concepts in diagnosis and treatment. *Journal of Clinical Gastroenterology*, 40(1):3–14, January 2006.
- [8] K. Zuiderveld. Contrast limited adaptive histogram equalization. *Graphics Gems IV*, pages 474–485, 1994.
- [9] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2002.
- [10] M. Häfner, L. Brunauer, H. Payer, R. Resch, F. Wrba, A. Gangl, A. Vecsei, and A. Uhl. Pit pattern classification of zoom-endoscopical colon images using DCT and FFT. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, Maribor, Slovenia, 20-22 June 2007.
- [11] M. Häfner, L. Brunauer, H. Payer, R. Resch, F. Wrba, A. Gangl, A. Vecsei, and A. Uhl. Pit pattern classification of zoom-endoscopical colon images using evolved fourier feature vectors. In *IEEE Workshop on Machine Learning for Signal Processing*, pages 99–104, Thessaloniki, Greece, 27-29 August 2007.
- [12] H. A. Mayer. Evolutionary algorithms in Java - JEvolution. <http://www.cosy.sbg.ac.at/~helmut>.
- [13] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

- [14] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [15] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.