Incorporating Human Knowledge in Automated Celiac Disease Diagnosis

Michael Gadermayr[1], Hubert Kogler[2], Maximilian Karla[2],
Andreas Vécsei[2], Andreas Uhl[3] and Dorit Merhof[1]

[1] Aachen Center for Biomedical Image Analysis, Visualization and Exploration (ACTIVE)
Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany
e-mail: Michael.Gadermayr@lfb.rwth-aachen.de

[2] St. Anna Children's Hospital, Department of Pediatrics, Medical University Vienna, Vienna, Austria

[3] Department of Computer Sciences, University of Salzburg, Salzburg, Austria

Abstract— **Recently, computer-aided celiac disease diagnosis has been promoted to provide an objective opinion besides histological examination of biopsies and visual assessment of macroscopic mucosal tissue. State-of-the-art techniques, however, are not accurate enough to provide incentive for clinical deployment. In this work, we answer two questions: Do computers and human experts make similar classification errors and can expert knowledge be utilized to increase the accuracy of computer-aided methods. Three experts were asked to perform visual classification of a large number of images. The experts decisions were combined with nine different state-of-the-art image representations. Experimentation showed that the correlations between two computer-based methods were higher than the correlations between an expert and a computer-based method. Furthermore, the inclusion of expert knowledge led to statistically significant ($p < 0.05$) improvements in 69 out of 108 investigated settings.**

I. Introduction

Celiac disease [27], [29], is a common autoimmune disorder primarily affecting the small bowel triggered by dietary gluten, which is the major protein found in various cereals. This disease is charaterized by an inflammation affecting the mucosa of the small intestine. During the course of the disease, the mucosa loses its absorptive villi and hyperplasia of the enteric crypts occurs. These mucosal alterations are associated with a diminished ability to absorb nutrients. The prevalence of celiac disease [10], [9], [3], [2], [8] varies geographically and across ethnic groups. In Europe and North America estimations range from 1:80 to 1:300.

Endoscopy with intestinal biopsies is currently considered the gold standard for the diagnosis of celiac disease. Microscopic changes found in these biopsies are then classified according to the modified Marsh classification [27], [29], which distinguishes between the classes Marsh-0 to Marsh-3, with subclasses Marsh-3A, Marsh-3B, and Marsh-3C. According to this classification, Marsh-0 signifies a healthy mucosa (without visible changes of the villous structure) and Marsh-3C designates a complete absence of villi (villous atrophy).

Previous studies on computer-aided celiac disease diagnosis [22], [16], [17], [24] considered the four classes Marsh-0 and Marsh-3A to Marsh-3C only, since visible changes of the architecture of the villi can be observed only for these classes. In this work, we focus on the two-classes case only (i.e. Marsh-0 vs. Marsh-3). The reason for working with this problem definition is given by the image data set available which is well balanced with respect to the images in each class only when using the two-classes case. Furthermore, this two classes case is most relevant for clinical practice.

A. Computer Aided Diagnosis

The current gold standard for detection of celiac disease is based on biopsies. This histological staging of biopsies is, however, subject to a significant intra- and inter-observer variability [35], [1], [28]. Therefore getting a second opinion can help to improve reliability (sensitivity/specificity) or to reduce the number of required biopsies [1]. Today, the histopathological findings are validated by visual assessment of the endoscopist. As the manual assessment of tissue by human expert's is again subject to high inter-observer variability [13], strong incentive is given to develop observer independent diagnostic methods such as a computer-aided diagnostic system.

Computer-aided diagnosis of celiac disease solely relies on image data captured during endoscopy of the duodenum (Fig. 1). Recently, significant research has been performed in the field of computer-aided celiac disease diagnosis. Classification has been done based on images obtained either by conventional endoscopy [22], [19], [20], [37], [14], [15] or based on wireless capsule endoscopy [32] [6], [4], [5]. All these methods focus either on obtaining the best possible overall classification rates based on a predefined data set [6], [4], [5], [22], [37], [20], or on selecting appropriate sub-images (patches) from original ones [14], [15]. These patches (Fig. 1) should ideally exhibit the specific disease markers (and should not suffer from strong image degradations) in order to achieve again the best accuracies utilizing a computer-based classification approach.

Although state-of-the-art image representations as well as machine learning methods have been applied, best accuracies are in the range of 85 to 90 % in the two-classes case. In authors opinion, a clinical deployment is currently inhibited by two factors:

1) Balanced accuracies between 85 and 90 % are not sufficient and
2) experienced medical doctors currently outperform computers-based methods in terms of accuracy [13].

978-1-4673-8910-5/16/$31.00 ©2016 IEEE

(a) Marsh-0

(b) Marsh-3

Fig. 1. Example image patches showing a healthy villi structure (a) as well as a diseased mucosa suffering from villous atrophy (b).

B. Contribution

In this work, we make use of physician's knowledge during interaction by asking them to visually classify the available image data. We investigate if computers and human experts make similar classification errors if characterizing endoscopic image data. Additionally, we assess if expert knowledge can be utilized to increase the accuracy of computer-aided methods. For that, we propose an approach which incorporates the decision of one medical expert into the final image representation. As the method relies on one medical expert only, it could be efficiently implemented in a clinical environment, which is of high practical relevance.

Recently [11], we already investigated a fusion of expert knowledge with computer-aided decision support techniques. In this work [11], which was highly clinical and thus application oriented, we investigated a small collection of computer-aided classification methods only and furthermore focused on the fusion of all available data available per patient to enhance the classification accuracy even further. Here, a significantly larger experimental setting is applied, investigating a large collection of image representations as well as combinations of them. Additionally, examinations on correlations among different features for automated classification and between features and expert rating are conducted in order to gain systematic insights into technical aspects of sensible fusion approaches.

II. Exploiting Expert's Decision

Theoretically, for each image, a decision can be computed with different computer-based classification methods. Additionally, in a clinical environment, the knowledge of the medical expert, who is performing the endoscopy, could be integrated into the further processing pipeline.

To gain insight and in order to identify sensible data for fusion, first the (linear) correlations between decisions of different computer-based methods as well as between computer-based methods and human experts are computed.

Furthermore, we propose a fusion approach to investigate whether any combination of computer- and human-based classification leads to improved overall classification rates. The focus in this scenario is on fusing computer-based methods with the knowledge on one human expert (as these two decision vector pairs are supposed to be less correlated). We specifically consider a scenario where only one physician provides exactly one binary decision per image. Theoretically, we could also incorporate data from several medical experts, however, such data would be hardly available in clinical practice. The major issue is how to fuse the available data [33]. The most intuitive strategy is to apply decision-level fusion. However, as we want to investigate the combination of one human expert with one computer-based technique, such a method would not work well as it requires more than two decisions per image to work effectively. Score-level fusion would demand from computer-based classifiers and human experts to provide soft decisions. Especially in the latter case this is not easy to obtain and highly unintuitive for physicians. Consequently, we decided for a special kind of feature-level fusion approach incorporating the decision of the expert directly into the feature vector of the computer-based method by means of concatenation. For that, after L^2 normalization of the feature vector and weighting of the expert's (binary) decision, the data is concatenated.

Finally, we additionally applied feature-level fusion to two computer-aided methods, with and without expert's information, to gain additional insight.

III. Experiments

A. Setup

The test data utilized for experimentation contains images of the duodenal bulb and the pars descendens taken during endoscopies at the St. Anna Children's Hospital. Prior to automated processing, all images are converted to gray scale images as the additional use of color information generally does not lead to consistent improvements [22], [37], [16], [17]. This conversion, however, is not applied for acquisition of experts' data, because human experts are used to color images and the rates could thereby suffer from a gray scale conversion.

In a pre-processing step, texture patches with a fixed size of 128×128 pixels were manually extracted to get more idealistic data (as positively evaluated previous research [16], [37], [22]).

To obtain the ground truth for the texture patches, the results of histological examination of biopsies from corresponding image regions are utilized. The staging of the villous atrophy was classified according to the modified Marsh classification [29]. A confirmation of the staging was obtained by visual inspection of an experienced endoscopist. Although it is theoretically possible to distinguish between several different stages of villous atrophy, in this study we aim at distinguishing between images from patients suffering from celiac disease (Marsh-3) and healthy patients (Marsh-0), as this two classes case is most relevant in practice.

Our experiments are based on three different balanced data sets, each containing 560 image patches (280 of class Marsh-0 and 280 of class Marsh-3). All overall (balanced) accuracies presented are based on the mean of 50 random splits. One distinct split divides the data set into an approximately balanced training (80 %) and evaluation set (20 %), restricting images of one patient to the same set in order to avoid any bias (due to similarities within data of one patient).

To gather decisions of human experts for a comparison, three experts in the field of gastrointestinal endoscopy manually annotated the image data. Thereby, 5040 manual classifications where performed (280 images $\times$ 2 classes $\times$ 3 configurations $\times$ 3 experts) providing a stable basis for further experimentation.

For the experiments, one decision per image was required in order to meet the practically most realistic scenario (as motivated in Sect. II) with only one available endoscopist during the examination. Therefore, the experimental evaluation is conducted individually for each human expert.

Data fusion is performed on feature-level. In case of fusing computer-based feature vectors with expert's decisions, the feature vector is concatenated with the decision label. Beforehand, the label is weighted (multiplied with a positive real valued scalar w) and the feature vector is L^2 normalized. The optimum weight w is evaluated (within $\{2^{-2}, 2^{-1}, 2^0, 2^1, 2^2\}$) based on the other two data sets (mean of the indexes is utilized).

In order to determine whether the performances of two techniques are statistically significantly different, the Mann-Whitney-Wilcoxon signed rank-sum test [26] is applied.

B. Feature Extraction Methods

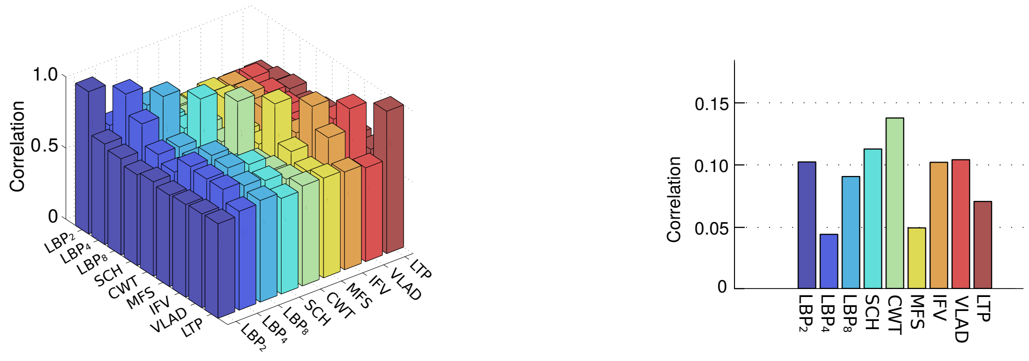
For a detailed comparison, several well known image representations are utilized. Beside state-of-the-art general purpose image representations [30], [36], [38], [34], [21], we investigate descriptors which have been specifically developed for analysis of endoscopic images [12], [23]:

- Local Binary Patterns [30] (LBP):
  LBP describe a texture by means of the joint distribution of pixel intensity differences represented by binary patterns. For experimentation, LBP is utilized with a radius of two pixels as well as various numbers of neighboring samples (2 (LBP_2), 4 (LBP_4) and 8 (LBP_8)) to gain insight into the impact of specific (not necessarily sensible) configurations.
- Local Ternary Patterns [36] (LTP):
  LTP is a generalization of LBP, aiming at a more robust final representation. This is achieved by introducing a different quantization scheme based on three states instead of the binarization applied in LBP. We utilize a standard configuration [36] (threshold $t = 5$, radius $r = 2$, eight circularly aligned neighbors).
- Multi-Fractal Spectrum [38] (MFS):
  This image descriptor is obtained by first computing the local fractal dimension for each pixel. This is performed using three different types of measures for computing the local density. Finally, the feature vector is built by concatenation of these fractal dimensions.
- Dual-Tree Complex Wavelet Transform [23] (DTCWT):
  This image descriptor is based on fitting a two-parameter Weibull distribution to the wavelet coefficient magnitudes of sub-bands obtained from the dual-tree variant of the complex wavelet transform. Decomposition is performed on five levels.
- Shape Curvature Histogram [12] (SCH):
  This method, which has been specifically designed to deal with endoscopic image data, describes an image as the histogram of contour curvature values. The final representation is obtained by first selecting contour pixels (by means of edge detection), followed by curvature estimation, based on edge filter responses. Finally all curvatures in contour regions are collected into a histogram consisting of eight bins.
- Improved Fisher Vectors [34] (IFV):
  Fisher Vectors [31], as well as the next descriptor (VLAD), is a global mid-level image representation that is obtained by pooling local image descriptors. In case of Fisher Vectors, the Gaussian mixture model is utilized to construct a dictionary, based on a local descriptor. For this local descriptor, we make use of the SIFT (Scale-invariant Feature Transform) [25] feature. The final Fisher Vector contains information how the parameters of Gaussian mixture model have to be modified to better fit the data. This is done by concatenating the means and the covariance deviation vectors. We apply the improved Fisher Vectors version [34] (based on Hellinger's kernel and L^2 normalization).
- Vector of Locally Aggregated Descriptors [21] (VLAD):
  VLAD is a local pooling technique, similar to Fisher Vectors. In opposite to Fisher Vectors, VLAD does not store any second-order information. Furthermore it utilizes k-means clustering instead of a Gaussian mixture model to generate the feature vocabulary. The feature vectors finally store information of the difference between the cluster centers and the pooled local descriptors.

For the final classification we apply a linear support vector machine (SVM) which has often been utilized in previous work on computer-aided celiac disease diagnosis [14], [22], [18] and also generally in recent work on texture recognition [7]. To avoid bias, the c-value is optimized ($c \in \{2^0, 2^1, ..., 2^{11}, 2^{12}\}$) based on inner cross-validation.

C. Results and Discussion

To study, whether a fusion of expert's knowledge and computer-based methods could be beneficial, we first look at the linear correlations between the classification accuracies of computer-based methods and between a certain computer-based method and a human expert (Fig. 2).



(a) Correlations computer-based methods

(b) Correlations computer vs. expert

Fig. 2. Correlations of classification accuracies between different computer-based techniques (a) and between computer-based techniques and an expert (b) in case of data set WL1.

Considering the correlations between two different computer-based methods (Fig. 2 (a)), we notice that the correlations are generally high (consistently above 0.5). If looking at the correlations between computer-based methods and a human expert (Fig. 2 (b)), it can be clearly seen that these correlations are distinctly lower (between 0.04 and 0.14). This provides evidence that a fusion of computer-based and human experts data could be advantageous whereas a fusion of two (more correlated) computer-based techniques is supposed to be less effective.

Figure 3 (a) - (i) shows the outcomes of combining a computer-based method with human expert's decisions. The bars indicate accuracies for computer-based classification (left), human expert classification (right) and the fusion (wide center bar). On top, the p-values of the significance tests are given (null-hypothesis: accuracies with (computer and expert) and without fusion (only computer) are identically distributed). A p-value less than $\alpha = 0.05$ typically indicates a significant improvement. Looking at these results, we notice that the fusion in general is definitely advantageous. In 99 out of 108 cases, the accuracy of the more appropriate approach (computer or expert) can be boosted applying the fusion-based approach. Furthermore, in 69 (with $\alpha = 0.05$) or 54 cases (with $\alpha = 0.01$) out of 108 cases, the (only) computer-based technique is statistically significantly outperformed. The presented results provide strong evidence that fusion of human expert's opinion with computer-based image descriptors is beneficial.

After these general considerations, we focus on the impact of the computer-based method on the overall (fusion-based) accuracy. We notice that the best overall performances are obtained in combination with LBP_8, LTP and IFV which already exhibit best outcomes without fusion. However, even based on these high-performing methods, with feature-level fusion, consistent improvement are observed.

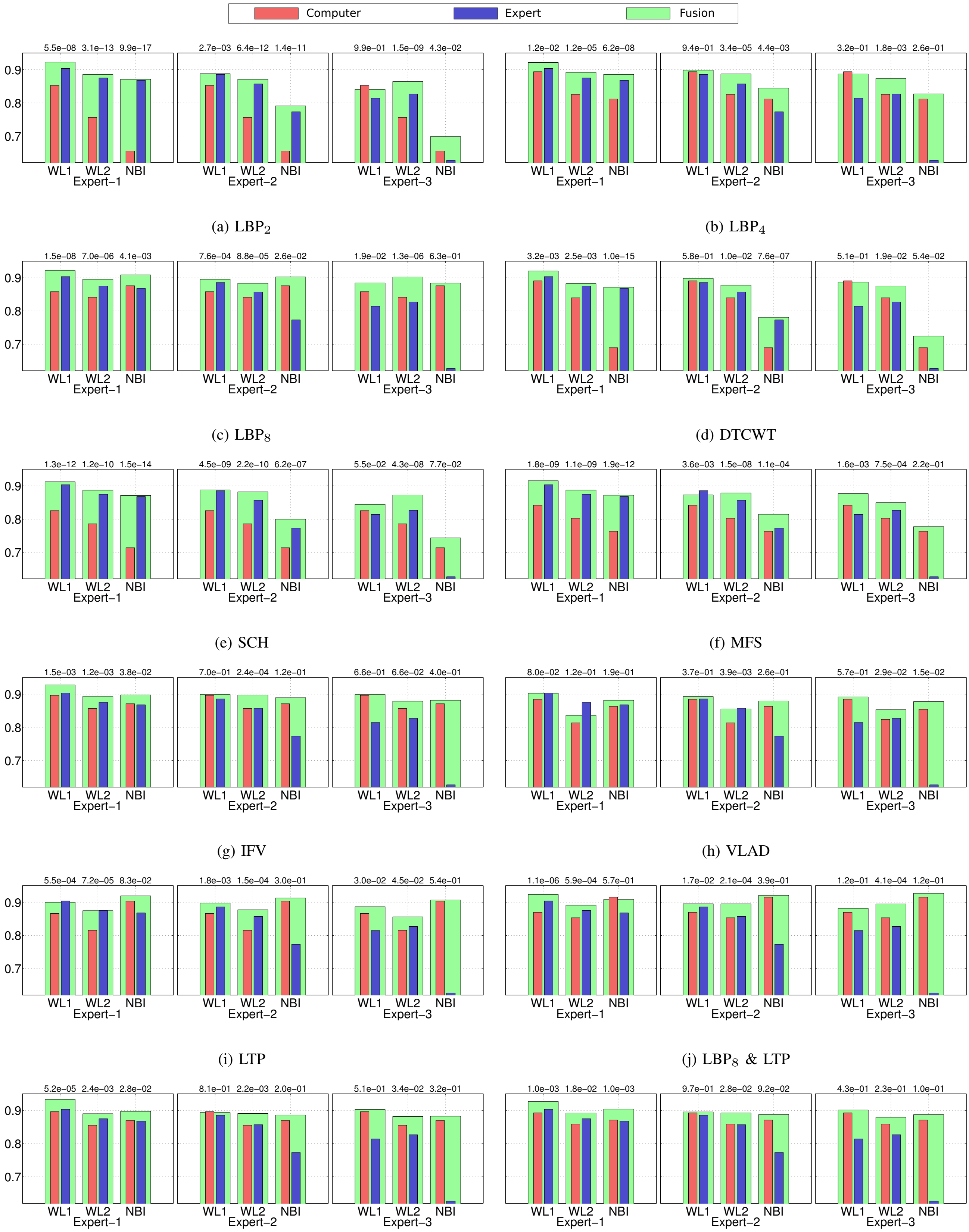
Finally, for the purpose of comparison, the impact of fusing two (more correlated) computer-based techniques (with and without expert's information) is explored as shown in Fig. 3 (j) - (l). In these plots, the left narrow bars indicate the accuracy of fusing two computer-based methods while the right bars present the rates of the human expert. The wide bars show the accuracies if fusing all data (two computer-based methods and the expert's data). In this evaluation focus is on all possible combination of the three best performing methods (LBP_8, LTP, IFV). We notice, that in general two computer-based methods do not continuously outperform the better one of the two approaches (as can be seen if comparing the overall results with Fig. 3 (c), 3 (g) and 3 (i)). This is supposed to be due to the higher degree of correlation between two computer-based methods.

IV. Conclusion

We investigated the fusion of computer-based methods with expert's knowledge. Focus was specifically on a scenario with one single expert only, allowing for an efficient clinical deployment without requiring any further manpower. It was shown that the overall classification rates (i.e. balanced accuracies) of computer-aided diagnosis can be boosted statistically significantly by adding human knowledge to the final image representation, in a feature-level-fusion sense. The proposed fusion-based technique was never significantly outperformed, even by experienced medical doctors. However, especially if an expert's accuracy was rather low, the classification performance is distinctly boosted. Based on quite small data (patches with 128×128 pixels) we obtained up to 93 % classification accuracy. Future work should be on an automated exploitation of the massive amounts of data available in endoscopic video material, which is likely to enhance the classification process even further.

Funding

This work has been funded by the Austria Science Fund (FWF) under project number KLI 429.



(a) LBP_2 (b) LBP_4

(c) LBP_8 (d) DTCWT

(e) SCH (f) MFS

(g) IFV (h) VLAD

(i) LTP (j) LBP_8 & LTP

(k) LBP_8 & IFV (l) LTP & IFV

Fig. 3. Classification accuracies obtained with fusion of computer-based data with human expert's decisions. The bars indicate accuracies for computer-based classification (left), expert classification (right) and the fusion (wide center bar). On top, the p-values of the significance tests are given (null-hypothesis: accuracies with (computer and expert) and without fusion (only computer) are identically distributed).

REFERENCES

[1] Arguelles-Grande, C., Tennyson, C., Lewis, S., Green, P., Bhagat, G.: Variability in small bowel histopathology reporting between different pathology practice settings: impact on the diagnosis of coeliac disease. J. Clin. Pathol. 3(65), 242–247 (2012)

[2] Castano, L., Blarduni, E., Ortiz, L., Nunez, J., Bilbao, J., I., R.: Prospective population screening for celiac disease: high prevalence in the first 3 years of life. J Pediatr Gastroenterol Nutr 39(1), 80–84 (2004)

[3] Choi, J., Lebwohl, B., Wang, J., Lee, S., Murray, J., Sauer, M.: Increased prevalence of celiac disease in patients with unexplained infertility in the united states. J Reprod Med 56(5–6), 199–203 (2011)

[4] Ciaccio, E.J., Tennyson, C.A., Bhagat, G., K., L.S., H., G.P.: Implementation of a polling protocol for predicting celiac disease in videocapsule analysis. World J Gastrointest Endosc 5, 313–322 (2013)

[5] Ciaccio, E.J., Tennyson, C.A., Bhagat, G., K., L.S., H., G.P.: Use of basis images for detection and classification of celiac disease. Bio-Medical Materials and Engineering (2014)

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

[7] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14). pp. 3606–3613 (2014)

[8] Dube, C., Rostom, A., Sy, R., Cranney, A., Saloojee, N., C., G.: The prevalence of celiac disease in average-risk and at-risk western european populations: a systematic review. Gastroenterology 128, 57–67 (2005)

[9] Dydensborg, S., Toftedal, P., Biaggi, M., Lillevang, S., Hansen, D., S., H.: Increasing prevalence of coeliac disease in denmark: a linkage study combining national registries. Acta Paediatr 101(2), 179–194 (2012)

[10] Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R.B., Drago, S., Elitsur, Y., Green, P.H.R., Guandalini, S., Hill, I.D., Pietzak, M., Ventura, A., Thorpe, M., Kryszak, D., Fornaroli, F., Wasserman, S.S., Murray, J.A., Horvath, K.: Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study. Archives of internal medicine 163, 286–92 (Feb 2003)

[11] Gadermayr, M., Kogler, H., Karla, M., Merhof, D., Uhl, A., Vcsei, A.: Computer-aided texture analysis combined with experts knowledge: Improving endoscopic celiac disease diagnosis. World Journal of Gastroenterology (WJG) (2016)

[12] Gadermayr, M., Liedlgruber, M., Uhl, A., Vécsei, A.: Shape curvature histogram: A shape feature for celiac disease diagnosis. In: Medical Computer Vision. Large Data in Medical Imaging (Proceedings of the 3rd International MICCAI - MCV Workshop 2013). Springer LNCS, vol. 8331, pp. 175–184 (2014)

[13] Gadermayr, M., Uhl, A., Vécsei, A.: The effect of endoscopic lens distortion correction on physicians' diagnosis performance. In: Proceedings of Bildverarbeitung für die Medizin 2014 (BVM'14). pp. 174–179. Springer Informatik aktuell (Mar 2014)

[14] Gadermayr, M., Uhl, A., Vécsei, A.: Getting one step closer to fully automatized celiac disease diagnosis. In: Proceedings of the 4th IEEE International Conference on Image Processing Theory, Tools and Applications 2014 (IPTA'14). pp. 13–17 (Oct 2014)

[15] Gadermayr, M., Uhl, A., Vécsei, A.: Quality based information fusion in fully automatized celiac disease diagnosis. In: Proceedings of the German Conference on Pattern Recognition (GCPR'14). Springer LNCS, vol. 8753, pp. 1–12 (2014)

[16] Gschwandtner, M., Liedlgruber, M., Uhl, A., Vécsei, A.: 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) (Nov 2010)

[17] Gschwandtner, M., Hämmerle-Uhl, J., Höller, Y., Liedlgruber, M., Uhl, A., Vécsei, A.: Improved endoscope distortion correction does not necessarily enhance mucosa-classification based medical decision support systems. In: Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP'12). pp. 158–163 (Sep 2012)

[18] Hegenbart, S., Uhl, A.: A scale- and orientation-adaptive extension of local binary patterns for texture classification. Pattern Recognition 48(8), 2633–2644 (2015)

[19] Hegenbart, S., Uhl, A., Vécsei, A.: Survey on computer aided decision support for diagnosis of celiac disease. Computers in Biology and Medicine (2015), accepted

[20] Hegenbart, S., Uhl, A., Vécsei, A., Wimmer, G.: Scale invariant texture descriptors for classifying celiac disease. Medical Image Analysis 17(4), 458 – 474 (2013)

[21] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34(9), 1704–1716 (2012)

[22] Kwitt, R., Hegenbart, S., Rasiwasia, N., Vécsei, A., Uhl, A.: Do we need annotation experts? a case study in celiac disease classification. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'14). Springer LNCS, vol. 8674, pp. 454–461 (Sep 2014)

[23] Kwitt, R., Uhl, A.: 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). pp. 1–8 (2007)

[24] Liedlgruber, M., Uhl, A., Vécsei, A.: 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) (Jul 2011)

[25] Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision (CVPR'99). vol. 2, pp. 1150 – 1157 (1999)

[26] Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18(1), 50–60 (1947)

[27] Marsh, M.: 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)

[28] Mubarak, A., Nikkels, P., Houwen, R., Ten, K.F.: Reproducibility of the histological diagnosis of celiac disease. Scand. J. Gastroenterol. 46(9), 1065–1073 (2011)

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

[30] Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (Jul 2002)

[31] Perronnin, F., Dance, C.: Fisher kernels on visual vocabularies for image categorization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07). pp. 1–8 (Jun 2007)

[32] Petroniene, R., Dubcenco, E., Baker, J.: Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. Am. J. Gastroenterol. 100(3), 685–694 (2005)

[33] Ross, A., Jain, A.: Information fusion in biometrics. Pattern Recognition Letters 24(13), 2115–2125 (2003)

[34] Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.J.: Image classification with the fisher vector: Theory and practice. International Journal of Computer Vision (IJCV) 105(3), 222–245 (2013)

[35] Taavela, J., Koskinen, O., Huhtala, H., Lähdeaho, M., Popp, A., Laurila, K., Collin, P., Kaukinen, K., Kurppa, K., Mäki, M.: Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. PLoS One 8(10) (2013)

[36] Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Analysis and Modelling of Faces and Gestures. Springer LNCS, vol. 4778, pp. 168–182 (Oct 2007)

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

[38] Xu, Y., Ji, H., Fermüller, C.: Viewpoint invariant texture description using fractal analysis. International Journal of Computer Vision 83(1), 85–100 (2009)